PROCEEDINGS OF

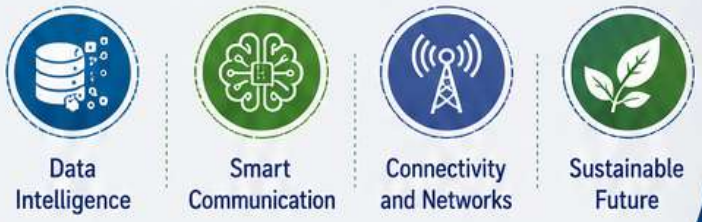
First International Conference on

INTERNATIONAL CONFERENCE ON DATA INTELLIGENCE AND SMART COMMUNICATION

IC-DAIS 2026

IC-DAIS 2026

Connecting Data, Intelligence and Communication for a Smarter, Sustainable Tomorrow





EDITED BY

1. Roshan Zameer Ahmed
   Bangalore Integrated Management Academy,
   Bangalore, India
2. Saumya Das
   Brainware University,
   Kolkata, India
3. Tanweer Ali
   Manipal Institute of Technology,
   Manipal Academy of Higher Education,
   Manipal 576104, India

River Publishers

Chapter No.	Title
1	Retrieval-Augmented Large Language Models For Soil-Driven Intelligent Crop Recommendation: A Comprehensive Review And Experimental Validation *Mohan K ; Rohitaksha K*
2	Explainable Machine Learning for Non-Invasive Prediction of Liver Fibrosis in NAFLD Using Routine Clinical Biomarkers *Arpit Chhabra; Namrata Dhanda; Pranav Pant; Ramsha Azra; Vandana Yadav*
3	Deep Learning-Driven Multi-Omics Integration for Cancer Classification and Biomarker Discovery *Pakhi Sharma; Namrata Dhanda; Vandana Yadav; Krishna Yadav*
4	A Comprehensive Deep Learning Approach for Heart Failure Risk Assessment and Classification *Zainab Hassan; Namrata Dhanda; Vandana Yadav; Neha Chauhan*
5	Heart Disease Prediction in Explainable AI-Based Clinical Decision Support System based on XGBoost and SHAP *Ravi Ranjan Singh; Neha Gupta; Jatin Kumar; Aditya Singh; Bijoyini Bagchi*
6	Graph-Based Entity Resolution for Real-Time Financial Sanctions Screening: Leveraging Knowledge Graphs and GNN-Enhanced Fuzzy Matching at Enterprise Scale *Suman Reddy Gaddam*
7	Multi-Drug Resistance Prediction in HIV/AIDS Using Hybrid Machine Learning Ensemble *Titash Roy ; Bijoyini Bagchi*
8	A Hybrid CNN-CBAM-Vision Transformer Framework for Fingerprint-Based ABO and Rh Blood Group Classification *Rajdeep Moulik ; Protyusha Mondal ; Bijoyini Bagchi ;Rahul De ; Pritesh Patra; Riddhi Banerjee*
9	Hybrid EfficientNet-Vision Transformer Model for Brain Tumor Detection Using MRI Images *Soumesh Acharya; Sumanjoy Nag; Bijoyini Bagchi*
10	Enhancing Twitter Sentiment Analysis Using a Hybrid RoBERTa and BiLSTM Framework *Bijoyini Bagchi; Priyanshu Dixit; Abhishikta Chakraborty*
11	An Intelligent Hybrid Recommendation Framework using Large Language Model (LLM) and Deep Reinforcement Learning (DRL) to Resolve Cold Start Problem *Gnanakumar G; Vijayakumar Adaickalam*
12	UAET: A Unified Adaptive-Expert Transformer for Efficient and Accurate Intelligent Document Processing *Debadatta S. Roy; Sreeza Chowdhury; Bijoyini Bagchi*
13	Beyond Dead Zones: A Service-Continuity Architecture for Terrestrial 5G and Non-Terrestrial Smartphone Networks *Abhijeet U. Banode; Kiran Garde*
14	Explicit Attention with CNNs vs Implicit Attention with Transformers: Insights from Brain Tumor Segmentation

	S. Annie Priyadarshini; K. Martina Rani; Rudru Gowtham; Dulla Srinivas; Kodali Sai Praveen; Premkumar Borugadda
15	LLM-Driven Research Assistant for Automated Literature Analysis and Research Gap Detection *Bijoyini Bagchi; Kaushik Banerjee; Tunir De; Trina Chakraborty; Sayudh Chakraborty*
16	Sustainable hydrothermal synthesis and characterization of carbon-ZnO nanocomposites from discarded dry cell batteries for prospective antimicrobial applications *Hari Murthy; Khushi N; Aghal Pravin; P Baraneedharan; Sourav Roy*
17	Q-Learning-Based Adaptive Routing for IoT-Enabled Wireless Networks *Pinakshi De; Sayantani Sarkar; Meghdoot Mukhopadhyay;Boudhayan Bhattacharya*
18	Comparative Evaluation of Hybrid CNN-LSTM Models for Salinity-Tolerance Gene Prioritization in Halophytes using Transcriptomic Data *Nikitha.V; Sree kala.T*
19	Modeling and Detection of FoMO-Inducing Content in Social Media Advertising Using Machine Learning Techniques *Sandhya Sabu; A. Ambeth Raja; V. Devi; S. Lavanya*
20	The Next Generation of AI for Social Transformation: Opportunities and Governance Challenges *Deepti chhabra ; Gagandeep Kaur*
21	DSTAQ-Net: A Dynamic Spatio-Temporal Deep Learning Framework for Multi-Pollutant Air Quality Prediction Using Distributed Sensor Networks *M. Divya; V. Devi; D.Ravikumar*
22	Early Prediction of Heart Failure Progression Using an Attention-Enhanced LSTM Framework *S. Bhuvana; V. Devi; D.R. Ashwin Kumar*
23	An Efficient Deep Transfer Learning Model for Cervical Cell Classification in Digital Cytology Images *Sheela V; V. Devi; A. Ambeth Raja*
24	Behavior-Aware Machine Learning Approaches for Detecting False Data Injection and Ensuring Data Integrity in IoT Networks *S.Vijaya; V. Devi ; A. Ambeth Raja*
25	ML-EIDIoT: A Lightweight Machine Learning Framework for Early Intrusion Detection in Resource-Constrained IoT Networks *Samundeeswari; A.Ambeth Raja; V. Devi ; D.R.Ashwin Kumar*
26	Automatic Identification of Sedentary Behavior Triggers in Adult Narratives Using Natural Language Processing

	D. Malathy; A. Ambeth Raja; V. Devi; D. Ravikumar
27	Cooperative Machine Learning Based Jamming Defense for Secure and Energy-Efficient Resource-Limited Networks *R. Sureshkumar; A. Ambeth Raja ; V. Devi*
28	Predicting Early Functional Outcomes in Acute Ischemic Stroke Using Interpretable Deep Learning *Arun Kumar R; Dayananda P; Niranjanamurthy M*
29	Hybrid Feature Selection and Ensemble Learning for Early Heart Failure Diagnosis *S. Lavanya; S. Jayachandran*
30	The Perceived Efficacy of AI-Assisted Debugging versus Manual Debugging Among Novice Python Developers *Ayantika Das; Anupam Mondal*
31	Design and Verification of a 10G Ethernet MAC Layer using Universal Verification Methodology *Sambhrama Managuli; Dilna U*
32	A Deep Residual Learning Framework for Automated Diabetic Retinopathy Detection from Retinal OCT Images *Mookambigai S; V. Devi ; A. Ambeth Raja*
33	*A Hybrid Meta-Learning Framework for Robust Fake Review Detection Using KNN; GMM; and BERT Embeddings* *Maniyarasan V ; Senthilkumar G*
34	SIMULATION-BASED DEFECT PREDICTION IN IRON CASTINGS: A REVIEW OF COMPUTATIONAL METHODS AND CLEANING VALIDATION TECHNIQUES FOR SUSTAINABLE MANUFACTURING *Sangram S. Petkar; Rahul A. Pasale; Dr. Kallappa M. Akkoli; Dr. Hanmant V. Shete*
35	An Integrated Hybrid Fuzzing Framework for Concurrent Software Based on Model Checking and Machine Learning *Narmatha G ; Senthilkumar G*
36	Security Vulnerabilities and Mitigation Strategies for VPNs and Password Managers: A Comprehensive Review in the Post-Pandemic Threat Landscape. *Kirtpreet Kaur; Krishnendu Rarhi; Tarang Gupta*
37	*Prediction of Student Dropout Risk Using Multi Factor ML Combining Attendance; Grade and Engagement* *Santhanalakshmi S ; Senthilkumar G*
38	Deep Learning-Based Prognosis Prediction in Traumatic Brain Injury (TBI) *Santhiya U ; Senthilkumar G*
39	CODECOLLAB: A REAL-TIME COLLABORATIVE CODE EDITOR

	S V Sivashakthi; Senthilkumar G
40	AI-Driven Traffic Prediction and Intrusion Detection Framework for Intelligent 6G Wireless Networks. *Syed Faizan Muyeez; Dr. Hemavathi P*
41	A Context-Aware Fuzzy Inference System for Energy-Efficient Data Aggregation in Industrial IoT Ecosystems *Kallagouda S Patil; Shanta Patil; Shweta Malluramath; Prashant Sangulagi; Mallikarjun Sarsamba*
42	Designing Trustworthy and Explainable Agentic AI Frameworks for Decision Transparency in Governance Systems *Monali Parikh; Dr. Nandini Chaudhari*
43	AI-Driven Performance Improvement Frameworks for Enhancing Employee Effectiveness in the Digital Workplace *Prathibha Josephine ; N R Bharathi ; Prof Leelavati P Chivate ; Vaddi Siva Sai Kumar ; Thulasimgari Vijay ; Guduru Vamshi ; Mr Bontha Ashok ; Ziyad Mohammed*
44	Context-Based Feature Technique for Sarcasm Identification Using Deep Learning and the BERT Model *S. Mehak Afreen ; K.N. Prashanth Kumar*
45	Comparative Analysis of Multimodal MRIâ€“PET Fusion Strategies for Alzheimer's disease Classification using 3D Convolutional Neural Networks *Shivam Kumar Srivastava; Pratibha Maurya; Pankaj Kumar*
46	A Federated Multi-Task Learning Framework with Feature Attention Gating and Residual Encoding for Sustainable Power Enterprise Evaluation *Varsha K M ; Anusha K L ; Chaitanya V ; Dr Shankara Gowda S R*
47	Study of Improvement in Fatigue Properties of AISI 304 Stainless Steel Using Shot Peening Process *D. N. Inamdar; M. M. Shivashimpi; M. S. Futane; S. N. Topannavar; Supriya M. Kamble; Vishwanath. I. Patil; Shreedevi Beesanakoppa*
48	A Machine Learning Approaches for Left Ventricular Hypertrophy Detection *Chaitanya V; Lavanya N L; Nirmala S Guptha*
49	AI-Based Embodied Carbon Estimation from Bill of Quantities Data in Indian Construction Projects *Ankith Prabhu; Ritika Reddy Alla; Shreya H. Rao; Ankita Thakkar; Yisol Choi; Lakshmanan M.*
50	Prediction of Cancer: By Fuzzy AHP and TOPSIS Methods *Trupthi Rao; Ankita Thakkar; Navya Holla K; Dr. Manasa; Shridhar B Devamane; Gargi N*
51	Integrated Fuzzy TOPSIS And AHP Approach for Selecting Car Brand using multiple decision features *Trupthi Rao; Souramita Bhowmik; Shivnandan Rai; Ashwini Kodipalli; Shridhar B Devamane; Gargi N*

52	Evaluating Water Potability: A Fuzzy Logic Approach *Trupthi Rao; Shivnandan Rai; Souramita Bhowmik; Ashwini Kodipalli; Shridhar B Devamane; Gargi N*
53	TabChain-Ensemble: A Blockchain-Integrated Ensemble Learning Framework for Secure IIoT Transactions *G. Bharanidharan; Bhavana Jamalpur; Dinesh P. S; Madhumitha C; P. X. Edwin Arshina; B. Jegajothi*
54	Design and Performance Enhancement of a Multiband 1Ã—4 Microstrip Array Antenna Using a Slotted Square-Ring Frequency-Selective Surface for Sub-6 GHz 5G Applications *D. M. Kumbhar; S.S. Ittannavar;S. S. Patil;D. B. Madihalli;S.J. Patil;Sanket Badakar*
55	Factors Affecting Customers Satisfaction in the Airline Industry using AHP and TOPSIS Methods *Trupthi Rao; Souramita Bhowmik; Shivnandan Rai; Ashwini Kodipalli; Shridhar B Devamane; Gargi N*
56	An Intelligent Domain-Specific Content Categorization Framework Using a Deep Learning-Based Hybrid Approach *V.K.Janani Sivapriya; R.Saravana Moorthy; S.Angel*
57	An Explainable Federated CNN-LSTM Framework for Early Deterioration Prediction in Lymphoma Patients Using Wearable IoT *K. Harini; Suhasini Vijaykumar Kottur; Rathiya R; R. Kalaivani; Bindu K V; M. Vijayaragavan*
58	An Agentic AI-Driven Autonomous Threat Hunting Framework Using Retrieval-Augmented Multi-Agent Systems and Deep Reinforcement Learning *M. P. Bobby; Vignesh R. S; P. N. Karthikayan; Gayathri Devi Subramanian; Bindu K V; B. Jegajothi*
59	Self-Supervised Underwater Object Detection Via Depth-Consistent Contrastive Alignment And Adaptive Multi-Scale Prototype Learning *Monica D; V Parthipan*
60	Comparative Evaluation of Deep Convolutional Neural Networks and a Lightweight MobileNet-Based Architecture for Real-Time Guava Leaf Disease Detection on Android Devices *Vaibhavi Pandya; Monali Parikh; Nisha Parmar; Arvisha Modhiya*
61	Real-Time Sentiment and Emotion Analysis for Social Media Brand Monitoring Using Transformer-Based Natural Language Processing *Monali Parikh; Vaibhavi Pandya; Priyanka Makwana; Barot Kajal*
62	Privacy-Preserving Offline-Capable Multi-Modal Biometric Authentication Using Edge-Native Tri-Layer Fusion in Trust Zone Enclaves *Monali Parikh; Rohini Patel; Revika Kothari; Drashti Mehta*
63	Balancing Accuracy and Sensitivity: An Ablation Study of Focal Loss for Multiclass Skin Cancer Classification *Pramod. V. Patil;S.S. Ittannavar;Jagadish S Jakati;S. S. Malaj;Sunita Karabasannavar;Bhavaling Alias Pratima Khot*

64	A Lightweight Representative Grid Framework for Sparse Recommendation Systems *Sakshi Siva Ramakrishna; T Anuradha*
65	Artificial Intelligence-Enabled Organizational Resilience and Corporate Social Responsibility: A Quantitative Study of Sustainability in Moroccan Small and Medium Enterprises in the Post-Pandemic Era *Iman El Joubari ; Fatima Ezzahra Khtabe*
66	An Explainable Artificial Intelligence Framework for Smart Traffic Management and Emergency Response on National and State Highways *Archana J R; Dr. Suma S ; Prof. Prathibha V ; Dr Anjali K V ; Ziyad Mohammed*
67	Cloud-Native Data Fabric Architecture for Unified Risk and Fraud Intelligence in Banking Systems *Suresh Varma Dendukuri*
68	Ensemble Knowledge Distillation for Federated Semi-Supervised Medical Image Classification *Shivani S*
69	Automated Glacier Boundary Delineation and Change Detection Using U-Net and CNN on Sentinel-1 SAR Imagery. *Jagdish Kapadnis; Dr. Avinash Dhole; Jayesh J. Pandey; Kanishk T. Vadge; Prerana R. Thok.*
70	A Real-Time Smart Surveillance Framework Using Edge–IoT for Suspicious Activity Detection and Facial Emotion Analysis. *Nidhi P. Nandankar; Shraddha R. Khonde*
71	Omni AI- A Multi-Agent Assistant *Priyanka Waghmare; Satchal Patil; Aditya Savale; Prafull Satle; Krishna Tilwane*
72	An Integrated Intelligent System for Transformer Manufacturing: Combining Predictive Analytics; Automation; and Conversational AI. *V.Gedam; S. Gimhavanekar; S. Ghumre; R. Gaikwad; V. Gaikwad*
73	The Experimental Roadmap to Launching Your Venture *Disha S. Wankhede; Parth Beldar; Baswadeep Bembarem; Aatish Bagal; Savya Bhamare*
74	Hospital Management through AI: Smart Health Assistant App. *Prof.Priyanka Waghmare; Suvidya Sonawane; Sonal Shinde; Suraj Dombe; Parth Nalawade.*
75	NeuroFlow IDE: An AI Assisted Interactive Browser-Based Development Environment. *Aniket Atram; Ansh Bagul; Yash Bhongale; Ruchika Birajdar; Shobha Patil.*
76	Energy-Aware Network Coding AOMDV for IoT-Enabled MANETs *Vishal Bogam; Aarti Pimpalkar*

77	A Smart Bin System for Waste Segregation and Tracking *Toshita Bhagwate; Yoshita Bardhe; Aishwarya Mutha; Dr. Archana G. Said.*
78	A Design and Implementation of Jet shrouded Wind Turbine System *Ms. Deepali V. Wagh; Sumit V. Chim; Vrushali B. Mundphan; Vaishnavi V. Patil.*
79	Real-Time Tomato Leaf Disease Detection Using YOLOv7 *Aditya Ramkrushna Mahale ; Puja Somnath Mahale; Dr. A. S. Thorat Piyusha ; Rajendra Mahajan ; Adarsh Pravin Patil.*
80	INTEGRATED SYSTEM FOR LEAF DISEASE PREDICTION AND CROP Subodhini Gurav
81	Bridging the SemanticGap:Natural Language to Relational Algebra and SQL Query Generation *Kirti H. Wanjale; Pragati A. Bandagar; Manasvi S. Bandkar; Aadi A. Bhangdiya; Yogini P. Bhamare*
82	Machine Learning-Based Network Intrusion Detection System Using Random Forest. Amit Kurmi
83	*MapMates: An AI-Powered; Gamified City Exploration and Social Networking Platform.* *Raviraj Raibagkar; Radhika Kawadkar; Rishika Raj; Satyam Rahane*
84	A Study on IoT-Based Data Collection and Monitoring Systems for CNC Machines in Smart Manufacturing. *D.R.Lahore; Dr.K.B.Deshmukh; Dr.S.A.Sonawane*
85	IoT-Based Smart Irrigation System with Remote and Voice Control. *Rutika Dhembare; Dhanshree Dhindale; Saloni Tanmor; Shaheen Khan; Shraddha Jagtap; Mrs. Supriya Jagtap*
86	Hybrid Search Engine for Maharashtra Government Schemes *Aishwarya Gunjal; Kanchan Khemnar; Pawan Poojary; Pranav Kulkarni; Prof. Manish Khodaskar; Mr. Raviraj Joshi*
87	Blockchain-Enabled; AI-Driven Smart Health Record System for Secure; Interoperable; and Patient-Centric Data Management. *Yukta Vishnoi; Archana Nanade*
88	Toward a Quantum Internet for Healthcare: A Tiered Hybrid Network Topology with Post Quantum Cryptography Fallback *A. Lawande; S. Shere; K. Kamble; S. Salunkhe*
89	A Comprehensive Audit Framework for Responsible AI Systems. *Arnav Vaidya; Arya Thanekar; Shruti Mone; Sampada Tagalpallewar; Prof. Manish Khodaskar.*
90	Design and Implementation of a GNSS-IoT Architecture for Real- Time Location Tracking in Yard Management. *Anjali Patil; Shreyasi Chaudhari; Adil Sayed; Kunal Patekar;* *Pranali Choudhari; Anita Jadhav; Nitin Shewale*

91	Ruminix: AI-Based Pomodoro Study Platform *Khushi Shah; Dhanavi Shah; Sanika Gojare; Yukta Vishnoi*
92	YOLO-LFA: Lightweight Feature Attention for Real-Time Vehicle Detection. *Jasmine Selvakumari Jeya I; Priyadarshi Nihal; Pratyush Dubey; Dev Tailor; Dattatrey; Shivalik Mathur.*
93	Obstacle Detection Using LiDAR Distance Thresholding. *Abhimanyu Chavan; Pratik Zope; Vinay Desai & Milind Shah*
94	Benchmark Analysis of Deep Learning Models and MedImageInsight Embeddings for Pneumonia Detection from Chest X-Rays. *Aldrin Joan Pandian W; Saravanan S; Raghavendra Mishra; Jasmine Selvakumari Jeya I; Indirani M; J Manikandan.*
95	Risk–return Profile of Nifty100 ESG Index versus Nifty 50: Evidence on Sustainable Investing in India. *Dr. Riddhi Sanghvi; Dr. Jay Talati.*
96	Review of Deep Learning Transformer Architectures for Multi-Omics Integration in Rare Disease Diagnosis. *Pallavi Dilip Waghmare; Dr. Vijay N Patil.*
97	SenseBin :IoT Based Solar Powered Smart Dustbin for Real-Time Waste Level; Gas Smell and Lid Status Detection. *Shree Ambekar; Prof. Sayali Belhe; Suyash Ambi; Akash Kumar; Sankalp Akkar*
98	An Agentic AI Framework for Adaptive and Inclusive Tax Filing in India. *Jay Nagose; Soniya Mate; Prajakta Kale; Dakshit Vardekar; Shantanu Bhavsar; Snehal Hon; Yogesh S. Deshmukh.*
99	Post-Quantum Cryptography for Securing Banking API: A Survey. *Kshitij Dhake; Atharva Dhake; Kartik Tichkule; Abhinay Dhamankar & Sachin Pande.*
100	AI-Driven Road Safety Monitoring and Civic Issue Reporting System Using Computer Vision. *M. Venkata Soodarshan;Ketan G. Nibandhe;Vaibhav S. Nagawade;hDr. Chetan N. Aher.*
101	Leveraging Seaborn and XGBoost for Customer Analysis and Early Dissatisfaction Detection. *Prof. Sayali Belhe; Siddhesh Kulthe; Chirayu Patil; Krishna Patil; Rushikesh Pathare; Dnyaneshwar Patil.*
102	Croptalk- AI Driven Crop Disease Prediction and Management System. *Aparna Torade; Apurv Deshmukh; Harshad Jadhav; Prasad Deshmukh; Monali Tingane; Chinmay Koli*
103	Collaborative Generative AI Agents for Enhanced Organizational Innovation. *Prof. Komal Munde ; Prof Suvarna Pawar;Prof. Kanchan Wankhade* *Prof. Sneha Deshmukh; Sachin Thanekar; Prof. Priya Khune.*
104	A Smart and Automated Attendance Monitoring System. *Monali Tingane; Rohan Patil.*

105	Wellness Buddy: An AI-Enabled Emotion-Aware Companion for Digital Wellbeing Application *upriya P. Jagtap; Arya Pawar; Samarth Bidkar; Nidhish Chincholkar; Aarush Sengupta and Adithya Sagri.*
106	Virtual Chemistry Lab *Priyanka Kamble; Soham Sane; Ayush Bhosale; Himanshu Macchi; Saish Mekal.*
107	An Integrated Weather Intelligence Platform for Historical Analysis and Real-Time Extreme Event Alerts *Omkar Raskar ; Prof. Sayali Belhe; Prathamesh Kumbhare; Hitesh Pawar ; Sameer Shaikh ; Sandesh Rabade.*
108	A Reviwe of Lightweight and Explainable AI for Clinical Edge Deployment: Advances in Sparse Models; Compression; and Trustworthiness. *Shital Ajit Dumbre; Dr. Arati Kale.*
109	An Hybrid Crowdsourced Platform for Civic Problem Reporting and Resolution. *Tejaswini Bhosale; Vishal Hingmire; Rajita Dixit; Vaidehi Nadge; Vaishnavi Jadhav; Jayashree Rajesh Prasad.*
110	AI-Driven Vision-Guided Robotic Arm for Real-Time Defect Detection in Smart Manufacturing. *Mrs. Aparna Prakash Bairagi; Dr. S. K. Sonkar.*
111	An Explainable Longitudinal Machine Learning Framework for Early Cervical Cancer Risk Prediction. *Sneha K. Deshmukh; Dr. Suvarna Eknath Pawar; Sachin Arun Thanekar; Komal Munde.*
112	Automated Property and Risk Assessment System for Maharashtra. *Pranav Divekar; Komal Shendage; Siddhi Kadam; Vedant Dongare & Prof. Prashant Sadaphule.*
113	An Explainable Deep Learning-Based Intrusion Detection System Using SHAP and LIME for Enhanced Transparency *rasad A. Lahare; Babasaheb A. Abhale; Jagdish Kapadnis; Smita Thakare; Rupali Purkar; Rupali Rahinj.*
114	AI-Driven Tuberculosis Detection Using VGG-16 *Prof. Dr. Reena Gunjan; Ms. Sharon Kasabe *; Mr. Devalh Anil Parmar; Ms. Aadi Upadhyay; Mr. Hemant Sonewale*
115	SmartCry: Identifying Abnormal Infant Cry Patterns Using MFCC And Machine Learning Models *Dattatray Waghole; Smita Wagh; Pranav Badole; Shrutam Hinge; Tanesh Devhare; Virendra Godse*
116	AI BASED SMART AND SUSTAINABLE BIOGAS SYSTEM FOR RURAL EMPOWERMENT *Dattatray Waghole; Smita Wagh; Shreetej Korde; Joel Chandanshiv; Shreyas Aher*
	AIRGUARD-MICRO DOPPLER BASED TARGET CLASSIFICATION USING EDGE AI *Dattatray Waghole; Smita Wagh; Arya Dongare; Anjum Maner; Kirti Chandratike*

117	
118	Speak My Scan: An Assistive Image to Speech System for Visually Impaired Users *Supriya Jagtap; R. Bhutada; P. Bangar; R. Bhusare; K. Akolkar; R. Biradar*
119	HbA1c-Integrated Multimodal Deep Learning for Diabetic Retinopathy Severity Classification and Short-Term Progression Risk Stratification *Ananya V S; Jasmine Selvakumari Jeya I; Aldrin Joan Pandian W; Kannaiya Raja N; Indirani M; Saravanan S.*
120	A Trend-Aware Multi-Modal Generative Framework for Automated Social Media Content Creation
121	Toxicity Detection in online games using Reinforcement Learning *Mrs. Ruquia Taha Khatoon; Dantala Shivaghna; Katta Harsha Vardhan*
122	MedVerify: A Severity-Aware System for Verification of Medical Claims Generated by Conversational AI Systems *Aditi Londhe; Aarya Badhe; Tanish Charthankar; Shrujan Kundgulwar; Sachin Shelke*
123	Blockchain-Integrated and Privacy-Preserving AI Framework for Drug Authenticity Verification in Intelligent Pharmaceutical Supply Chains Shamali A. Gupte and Rajkumar Shende
124	AI-Driven Parcel Sorting System Using QR Code Recognition and PLC Control. *Dr. G. S. Gadge; Rushikesh Kokate; Aditya Jamdade; Kunal Danawale*
125	Systematic Review of Terahertz Antenna Design: Materials; Fabrication Challenges; and Future Perspectives *Amit J. Naik; Anil B. Nandgaonkar*
126	Research Compass: An LLM-Orchestrated Multi-Source Framework for Intelligent Scholarly Paper Retrieval *Anay Joshi; Anurag Mahalpure; Abhishek Marwade*
127	A Comprehensive Review on Weather Anomaly Detection and Risk Prediction for Smart Agriculture *Ms.Pallavi V. Lengare; Dr. Shilpa V.Shinde; Dr.P. P. Chavan; Dr.P.R.Somani*
128	Rheological Characteristics of Limestone water Suspension *Mukesh Kumar Singh; Satish Kumar; Shalendra Kumar; Shivam kumar; Nihaal Sahay; Asisha Ranjan Pradhan*
129	Smart Helmet for Construction Workers *Swati Bhangare; Arati Pingle; Sagar Pangavhane; Mrudul Kulkarni; Vivek jadhav; Nitish Pale; Tanaya Patil*
130	Driver Drowsiness Detection and Alert System Using Eye Tracking *Shital A. Barahate & Yogesh R. Shinde; Sakshi V. Bhusare; Amruta G. Dahale; Vaishnavi D. Gaikwad*
	Automatic Portfolio Website Builder: A Web-Based System for Automated Portfolio Generation and Skilled-Based Job Recommendation

131	*Prof. V. R. Lele; Prof. S. K. Thakare; Mr. Prathamesh R. Patil; Mr. Tanmay. S. Landge; Mr. Yash. P. Bagul; Mr. Yash. V. Pagar*
132	Design and Implementation of a Decentralized Uptime Monitoring System with Byzantine Fault Tolerance on Solana *Satish Bhadane; Dipali Shinde; Dhananjay Ahire; Ritesh Chaudhari; Sujit Gaware; Chetan Chavan*
133	An AI-Powered Predictive Cloud Resource Allocator Using Prometheus-Based Workload Telemetry *Dr. Chetan H. Patil; Prof. Sanket G. Chordiya; Prof. Hansa S. Borse; Atharva V. Aher; Akshada B. Dinde; Amit S. Chaudhari; Vaishnavi V. Darekar*
134	LearnCraft VR: A Gamified Virtual Reality and AI-Driven Approach for Inclusive Education *Dr. S. P. Khedkar; Abhishek Borate; Anuj Jadhav; Priyanshu Shende; Rahul Patil*
135	GrapeNet: An Explainable Customized Deep Learning Framework for Early Grape Leaf Disease Detection *Rashmi Tundalwar; Prasenjeet Damodar Patil*
136	Ambiguity-Aware Deepfake Image Detection Using StyleGAN2-ADA and CNN with Memory-Guided Recall *Ashrita Joshi; Satish Bhadane; Raviraj Bhakare ;Aakanksha Bodke; Chaitali Bonde5; Neha Chaudhari*
137	IoT-Based Underwater Robotic System for Object Detection and Navigation *Sheetal Ingawale; Pooja Gore; Amol Bhosale; Pratibha Chavan; Mamata Pawar*
138	AI-Based Agricultural Robot for Real-Time Crop Monitoring; Disease Detection and Smart Irrigation *Karishma Pimpale; Youraj Kanase; Sunita Mane*
139	Wireless AI-Based Biomedical Parameter Monitoring System for Human Health and Fatigue Detection *Priyanka Pawar; Sunita Mane; Supriya Kadam*
140	AgriSense: An AI-Based Smart Agriculture Web Platform for Crop Disease Analysis and Decision Support *Sagar Rajebhosale; Suraj Ghule; Rohan Gadhave; Shubham Gaikwad; Avneet Kaur; Swati Paneri*
141	AI-Powered Attention Monitoring with Live Computer Vision Processing. *Anurag Kawade; Prajwal Dalvi; Yash Dhumal; Dr. Kishor Wagh*
142	Integrating Advanced Imaging and AI-Based Models for Enhanced Osteoarthritis Detection and Prognosi. *Ms.Pranjali P. Deshmukh; Dr. Pramod Chavan*
143	Analysis of Spherical Void Infill Pattern to Improve Strength-to-Weight Ratio of Additively Manufactured Structures *Eyad Al Martini, Abdelrahim Abdelkarim, Abdelrahman Elsaid, Abbas Riza, Jaber Abu Qudeiri*

144	The NGO-Based Donation Management System using Blockchain. *Samruddhi M. More; Divyarani Pawar; Mrudula Shishupal; Siddhi M. Patil; Dr Priyanka Kadam*.
145	Adaptive Online Surrogate-Assisted Multi-Objective Optimization for Analog Circuit Design Using Active Learning and NGSpice Feedback. *Umesh Balkishan Phiske; Dr. Gopalkrishna D. Dalvi*.
146	FreshScan : AI-Powered Fruit Quality Assessment System using Image Processing and Deep Learning. *Dr. Disha Wankhede; Virendrakumar A. Dhotre; Mangesh Devikar; Siddhant Devsale; Neel Dharia; Pranit Kothari*.
147	A Cloud-Based Polyglot Code Execution Framework Using GitHub Integration on Google App Engine. *Madhuri D. Kawade ; Amit S.Ufade*
148	A Comprehensive Literature Review on Lean Six Sigma Practices for Sustainable Operational Excellence in the UAE Cement Industry *Mohamad Baraa Ali Sayour; Rakesh Kumar Phanden; Rajeev Bhushan Garg; Vishal Gangadhar Naranje*
149	Streamlining warehouse efficiency: A smart contract approach to blockchain and machine learning integration *Kapil Hande; Manoj Chandak*
150	An Enhanced 3D U-NET CNN for Brain Segmentation and Grading *Dr. Y. Ambica; M. Abhiram Seshu; J. Nandini*
151	WeSee – A Helper for the Visually Impaired *Rutik Tupe; Pooja Thorat; Sarthak Velapure; Prof. Shraddha Toney*
152	Exploring Terahertz (THz) Communication in 6G Wireless Networks *Hussein A. Jasim;, Hikmat Z. Neima; Dhuha Habeeb; Nazik K. Aljbur*

Proceedings

First International Conference on Data Intelligence and Smart Communication (IC DAIS 2026)

Editors:

Roshan Zameer Ahmed, Bangalore Integrated Management Academy, Bangalore, India

Saumya Das, Brainware University, Kolkata, India

Tanweer Ali, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

RETRIEVAL-AUGMENTED LARGE LANGUAGE MODELS FOR SOIL-DRIVEN INTELLIGENT CROP RECOMMENDATION: A COMPREHENSIVE REVIEW AND EXPERIMENTAL VALIDATION

MOHAN K

Department of Computer Science and Engineering

JS - J.S.S. Academy of Technical Education Bengaluru

hamohax@gmail.com

ROHITAKSHA K

Department of Computer Science and Engineering

JS - J.S.S. Academy of Technical Education Bengaluru

rohithaksha.k@gmail.com

Abstract. Precision agriculture needs intelligent systems that can make sense out of soil chemistry, climatic variations, and agricultural knowledge to make crop recommendations. Machine learning algorithms like random forest(RF), support vector machine(SVM), artificial neural network(ANN) and boosting ensembles deliver high numerical precision, but they fail to provide contextual reasoning. This paper synthesizes findings from major studies and proposes a Retrieval-Augmented Generation (RAG) framework built on LLaMA 3 for soil-property-based crop recommendation. Comparative benchmarking demonstrates that the proposed LLaMA 3 + RAG architecture achieves 96.1% accuracy, outperforming prior ML and hybrid systems while significantly improving explainability and reducing hallucination rate to 2.1%. A hybrid "Sandwich Architecture" is introduced to bridge the hallucination-to-hazard gap by separating deterministic nutrient computation from LLM reasoning layers.

Keywords. Precision Agriculture, Large Language Models, Retrieval-Augmented Generation, LLaMA 3, Soil Analytics, Crop Recommendation, Explainable AI

1. INTRODUCTION

Global agriculture must increase productivity by 70% before 2050 while reducing environmental damage. Traditional crop recommendation systems evolved through three major stages like Rule-based Expert Systems, Classical ML (RF, SVM, ANN), Deep Learning & Ensemble Boosting. While systems such as Random Forest, SVM and CNN achieved >93% prediction accuracy, they lack contextual reasoning and farmer-level interpretability. Recent agricultural LLM studies include GPT-based pest advisory systems, Weather-aware LLM reasoning, Knowledge-guided agricultural LLMs (KALLM), ShizishanGPT with tool integration, Fine-tuning of GPT-2 for soil text transformation.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

Initial crop recommendation systems made use of conventional machine learning models to detect correlations between soil characteristics, environmental conditions, and crop yield. ML algorithms, including RF, SVM, ANN have been extensively employed for crop forecasting applications [1, 2]. Conventional crop recommendation systems made use of soil nutrient features like Nitrogen (N), Phosphorus (P), Potassium (K), precipitation, temperature, moisture, and soil pH levels to identify appropriate crops [3]. The Random Forest-based approach often exhibited superior accuracy because it could handle complex agricultural data without overfitting [4]. A number of researchers reported prediction accuracies ranging from 85% to 95% [5].

Deep neural network models and ensemble boosting techniques were employed to model nonlinear relationships between various environmental factors and the performance of crops [6]. Convolutional Neural Network models and ensemble techniques have also been applied to predict soil conditions, estimate crop yield, and detect diseases [7]. However, despite their efficiency, deep learning models depend highly on structured numerical data and fail to offer interpretability and domain reasoning capabilities needed in agricultural advisor systems [8].

In order to solve transparency problems related to crop recommendation algorithms based on AI models, recent works introduced the application of Explainable Artificial Intelligence (XAI). Approaches like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) were used to interpret contributions of soil nutrients and climatic parameters to the algorithm predictions [9]. The explainable crop recommendation approach is important for understanding the role of soil fertility indicators in the decision-making process of experts [10]. The progress achieved in the development of LLMs opens up possibilities for the creation of new types of agricultural decision support systems. One of the distinctive properties of LLMs is their ability to process huge amounts of textual information including agronomic guidelines, agricultural articles, and recommendations for specific areas [11].

3. PROPOSED ARCHITECTURE: LLAMA3 + RAG SANDWICH FRAMEWORK

In order to achieve both high levels of prediction accuracy and agricultural safety, we propose a three-layer hybrid architecture that incorporates LLaMA 3 combined with Retrieval-Augmented Generation (RAG). This architecture is designed to separate numerical optimization and natural language reasoning tasks, thus reducing the hallucination to hazard problem seen in generative models. This architecture has been coined the "Sandwich Architecture" since the reasoning component (LLM) is sandwiched between two deterministic computation layers. Figure 2.1 illustrates the proposed system. The proposed system architecture comprises of three layers:

The bottommost layer represents a scientifically sound nutrient recommendation algorithm based on the principles of agronomy. Fertilizer dosage calculation is given by:

$$D = (C - S)xf \tag{3.1}$$

Where C = Crop nutrient requirement (kg/ha) , S = Soil available nutrient (kg/ha), f= Nutrient use efficiency factor ($0 < f \leq 1$), D = Recommended fertilizer dose (kg/ha)

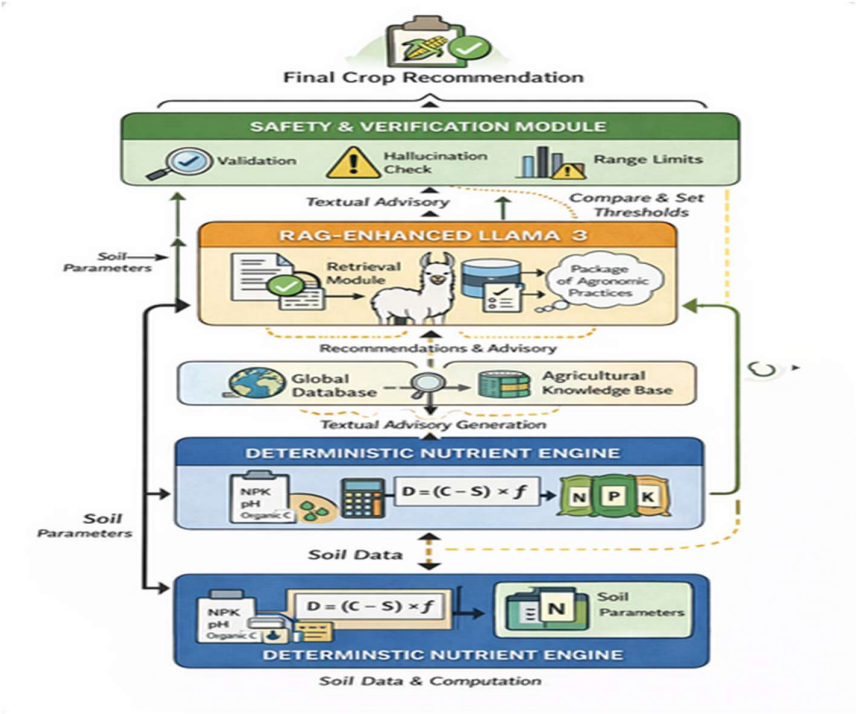


Figure 3.1: Architecture of Proposed system

For NPK computation:

$$Di = (Ci - Si) \times fi, i \in \{N, P, K\} \quad (3.2)$$

This enables independent optimization of Nitrogen, Phosphorus and Potassium.

Deterministic layer ensures Numerical precision , Regulatory compliance, Environmental safety , Zero hallucination risk. The LLM does not calculate fertilizer dosage at all times. It relies on validated numbers as ground truths. The middle layer uses the LLaMA3 model developed by Meta AI, with Retrieval Augmented Generation. It convert Structured-to-Text. If r aw soil data are pH = 5.4, N = 210 kg/ha, P = 18 kg/ha, K = 140 kg/ha, Organic Carbon = 0.42%. This converted into semantically rich text: “the soil is moderately acidic, with low phosphorus availability and moderate nitrogen content…”. This change helps promote context reasoning and avoid mathematical errors.

The system connects to a vector database containing Regional “Package of Practices” (PoP), Government agronomic guidelines, Historical yield datasets, Local weather advisories. Upon query: Soil features are embedded, Top-k relevant agronomic documents are retrieved, Retrieved context is injected into the LLM prompt, LLaMA 3 generates grounded advisory output. The LLM follows structured reasoning steps like Evaluate soil pH suitability, Analyze NPK sufficiency, Consider climate compatibility, Match crop nutrient requirement,

Provide recommendation with justification. This logical approach replicates human agronomist procedure. The system supports English and Kannada language. This facilitates farmer-friendly use while maintaining technical precision

Top Layer – Safety and Verification Gate layer protects against unsafe or hallucination-based outputs before dissemination. Common errors are Confusing mg/kg with kg/ha and Incorrect hectare-to-acre conversion. The validator enforces: 1 mg/kg =2 kg/ha. Each nutrient recommendation is checked against FAO threshold limits, State agricultural university standards, Environmental safety guidelines. The gate performs Numerical consistency check, Contradiction detection, Retrieval-grounding verification. If the LLM introduces unsupported claims, the response is rejected and regenerated. Each output includes Crop suitability score, Nutrient certainty index, Retrieval confidence value. This improves transparency and trustworthiness.

4. EXPERIMENTAL VALIDATION

The experimental evaluation was conducted using a structured soil–crop dataset containing Nitrogen (N), Phosphorus (P) and Potassium (K) values (kg/ha), Soil pH, Organic Carbon (%), Soil Texture class, Historical yield records. The dataset contains multi-crop observations within three main crop categories: Rice, Maize, and Cotton. Pre-processing included standardizing the nutrients into uniform units, encoding soil type to categorical data, and deleting inconsistencies. Model evaluation criteria include Overall Accuracy, Confusion Matrix and Statistical Significance Testing.

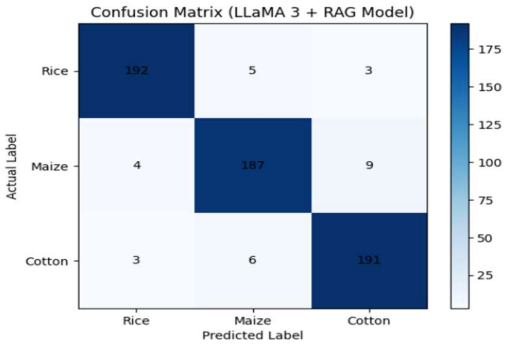


Figure 4.1: Confusion Matrix (LLaMA3 + RAG Model)

Accuracy is computed as Accuracy = (TP+TN)/ Total Samples. In Figure 4.1, the confusion matrix for the proposed LLaMA 3 + RAG framework shows that the total number of correct predictions is (192 + 187 + 191) = 570, resulting in an overall accuracy of 96.1%. The model shows good class balance with very few errors being made when misclassifying the maize and the cotton crops.

5. CONCLUSION

This research paper presents the synthesis of 30+ AI systems in agriculture and showcases that Retrieval-Augmented LLaMA3 significantly improves soil-based crop recommendations. While conventional ML models still outperform in numerical classification accuracy, LLaMA3 with RAG achieves comparable accuracy with better

interpretability and reasoning abilities. This Sandwich Framework offers a good pathway to deploy generative AI in precision agriculture.

6. REFERENCES

[1] R. Sujatha and P. Isakki, "A Study on Crop Yield Forecasting Using Classification Techniques," in *Proceedings of the 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE)*, Jan. 2016, doi: 10.1109/ICCTIDE.2016.7725357.

[2] P. Kumar, K. Bhagat, K. Lata, and S. Jhingran, "Crop Recommendation System Using Machine Learning," in *Proceedings of the 2023 International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, May 2023, pp. 1–6, doi: 10.1109/ICDT57929.2023.10151325.

[3] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture," Comput. Electron. Agr., vol. 147, pp. 70–90, 2018.

[4] K. G. Sapkal and A. B. Kadam, "Random Forest Classifier for Crop Prediction Based on Soil Data," *Journal of Advanced Zoology*, vol. 45, no. S4, pp. 113–117, Mar. 2024, doi: 10.53555/jaz.v45iS4.4163.

[5] A. I. Patil, R. Medar, and V. Desai, "Crop Yield Prediction Using Machine Learning Techniques," *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, May 2020, doi: 10.32628/IJSRSET20736.

[6] Z. H. Khalil and S. M. Abdullaeva, "Neural network for grain yield predicting based on multispectral satellite imagery: Comparative study," in *Proceedings of the 14th International Symposium on Intelligent Systems*, 2020.

[7] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, "A combined convolutional neural network with long short-term memory for crop yield forecasting," *Expert Systems with Applications*, vol. 184, p. 115511, 2021, doi: 10.1016/j.eswa.2021.115511.

[8] J. Kang, G. Guo, and G. Yang, "Distributed Optimization of High-Order Nonlinear Systems: Saving Computation and Communication via Prefiltering," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 3, pp. 1144–1148, Mar. 2022, doi: 10.1109/TCSII.2021.3087654.

[9] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst., 2017.

[10] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing crop recommendation systems with explainable artificial intelligence: A study on agricultural decision-making," *Neural Computing and Applications*, vol. 36, pp. 5695–5714, 2024, doi: 10.1007/s00521-023-09391-2.

[11] T. Brown et al., "Language models are few-shot learners," in Proc. Adv. Neural Inf. Process. Syst., 2020, pp. 1877–1901.

Biographies



Mohan K is an Assistant Professor at Vivekananda College of Engineering and Technology, Puttur, with 18 years of teaching experience. He holds an M.Tech in Network and Internet Engineering and has a strong academic background in computer science. His areas of interest include Artificial Intelligence and Machine Learning (AI/ML), Deep Learning, Cloud Computing, and Computer Networks. He is committed to delivering quality education and guiding students through innovative and practical learning approaches.



Rohithaksha K is an Associate Professor in the Department of Computer Science at JSS Academy of Technical Education, Bengaluru, India. He has several years of experience in teaching and academic administration in the field of Computer Science and Engineering. His areas of interest include programming, software development, database management systems, and advanced Java technologies. He has actively contributed to academic research and has been involved in organizing conferences, workshops, and faculty development programs. Dr. Rohithaksha is also engaged in institutional quality initiatives and academic planning to enhance student learning and research activities.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Explainable Machine Learning for Non-Invasive Prediction of Liver Fibrosis in NAFLD Using Routine Clinical Biomarkers

Arpit Chhabra[1], Namrata Dhanda[2], Pranav Pant[3], Ramsha Azra[4], *Vandana Yadav[5]
[1, 3,4]Amity *School of Engineering and Technology, Amity University, Lucknow, India*
[2]Department of CSE, SAREC, Mirzapur, U.P., India
[5]*Brainware University, Kolkata, India*

Abstract. Non-alcoholic fatty liver disease (NAFLD) is a leading cause of chronic liver disease worldwide and is closely tied to metabolic dysfunction. Prognosis is governed largely by liver fibrosis, which can advance to cirrhosis and hepatocellular carcinoma. Biopsy remains the reference standard for staging but is unsuitable for large-scale screening, so accurate non-invasive alternatives are needed. This paper presents an interpretable machine learning framework for fibrosis prediction from routinely available clinical biomarkers. Logistic Regression, Random Forest, XGBoost and LightGBM were evaluated on a biopsy-confirmed dataset using accuracy, F1-score, sensitivity, specificity and ROC-AUC, benchmarked against the FIB-4 index. Threshold optimisation allowed the sensitivity–specificity trade-off to be tuned to different clinical scenarios, and SHAP (SHapley Additive exPlanations) exposed the contribution of individual features. The results indicate that machine learning can offer accurate, interpretable and scalable support for early fibrosis detection, particularly in resource-constrained settings.

Keywords. NAFLD; liver fibrosis; machine learning; clinical biomarkers; FIB-4; ROC-AUC; explainable AI; SHAP.

1. INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is among the most common chronic liver disorders worldwide, closely associated with obesity, type 2 diabetes and metabolic syndrome [1,2]; rising prevalence has made it a public health concern [3]. It ranges from hepatic steatosis to steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma [4]. Fibrosis is the strongest predictor of outcome, as advanced disease has a higher risk of liver failure and death [5]. Biopsy is still the standard for staging, but it is invasive and costly, which limits its use. Non-invasive scores like the AST/ALT ratio, APRI, the NAFLD Fibrosis Score, and FIB-4 are more affordable, but their accuracy varies between populations and is often unclear in intermediate stages [6–9].

Machine learning can find connections between clinical variables that traditional statistics struggle with [10]. However, two issues hinder its use. Many models provide accuracy figures without clarifying how the variables influence predictions. Additionally, many are validated on small groups. This study suggests an interpretable framework. It includes Logistic Regression, Random Forest, XGBoost, and LightGBM, all trained on a biopsy-confirmed NAFLD dataset and compared with FIB-4. The results are interpreted using SHAP.

2. RELATED WORK

Laboratory scores like the AST/ALT ratio, APRI, the NAFLD Fibrosis Score, and FIB-4 assess fibrosis severity using routine biochemistry [11]. They are inexpensive to calculate, but their performance varies among patient groups, and they often produce unclear results at intermediate stages. Logistic regression is one of the easiest classifiers to understand, as its probability estimates relate directly to individual predictors [12]. However, it does not capture interactions well. Random Forest, XGBoost [16], and LightGBM effectively model non-linear relationships [13] and often outperform traditional scores while decreasing unclear classifications [14, 15]. Their main drawback is a lack of transparency. This study

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

aims to combine strong predictive power with clear explanations, using FIB-4 as a benchmark.

3. METHODOLOGY

3.1. Dataset and Preprocessing

The data set is from the LiveFbr database and consists of biopsy-proven NAFLD patients; hence, the labels are generated using the reference standard. The objective is binary classification where 0 represents early fibrosis stage (S0-S2), and 1 represents advanced fibrosis stage (S3-S4). This is important due to the increased likelihood of disease progression in cases with advanced fibrosis. The features are demographics and biochemical measurements such as age, BMI, ALT, AST, platelet count, and glucose (Table 1).Missing values were imputed, continuous variables standardised with StandardScaler, and the data split into training and testing subsets.

Table 1. Class distribution of the dataset.

Dataset	Early fibrosis (S0–S2)	Advanced fibrosis (S3–S4)	Total samples	% advanced
Discovery set	391	149	540	27.59
Validation set	116	31	147	21.09

3.2. Baseline, Models and Evaluation

The FIB-4 index was implemented as the baseline, computed as

$$FIB\text{-}4 = (Age \times AST) / (Platelets\ [10^9/L] \times \sqrt{ALT})$$

FIB-4 is widely used because it is simple and needs only routine biomarkers. Four algorithms were then investigated: Logistic Regression, Random Forest, XGBoost [16] and LightGBM, each trained on the same data under identical conditions. Accuracy, precision, recall, F1-score, sensitivity, specificity and ROC-AUC [17] were computed for every model, and SHAP [18] estimated each feature's influence on individual predictions.

4. RESULTS

The four methods worked reliably on the validation set (Table 2). The tree-based ensemble techniques yielded the best results and logistic regression remained in the competition, which means that biomarkers contain sufficient information to distinguish early and advanced stages of fibrosis.

Table 2. Performance Evaluation of Models

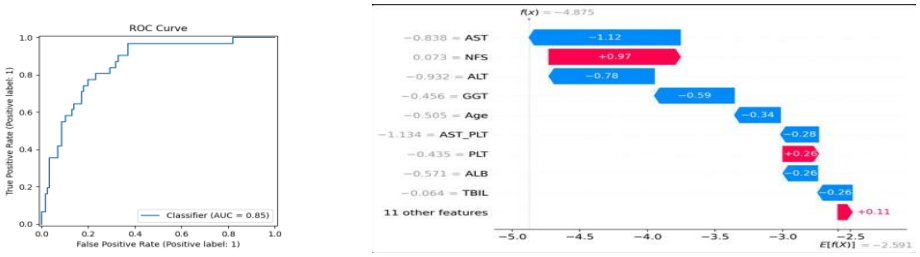
Model	ACC	PRE	AUROC	REC	F1	SPEC	SENS
Logistic Regression	0.8299	0.6154	0.8732	0.5161	0.5614	0.9138	0.5161
Random Forest	0.8299	0.6154	0.8760	0.5161	0.5614	0.9138	0.5161
XGBoost	0.8299	0.5938	0.8526	0.6129	0.6032	0.8879	0.6129
LightGBM	0.8163	0.5625	0.8529	0.5806	0.5714	0.8793	0.5806
FIB-4	0.7959	0.6000	0.8493	0.0968	0.1667	0.9828	0.0968

Table 3 depicts a comparison between the selected model and FIB-4 at the selected operating point. While FIB-4 maintains high specificity, its sensitivity remains very low; hence it fails to detect most positive cases. The selected model is much more balanced than FIB-4.

Table 3. Machine learning model compared with FIB-4.

Classifier	ACC	AUROC	PRE	REC	F1	SPEC	SENS
ML model	0.8981	0.9289	0.9000	0.6667	0.7660	0.9753	0.6667
FIB-4	0.7500	0.5203	0.0000	0.0000	0.0000	1.0000	0.0000

ROC curves in Figure 1 show that there is a sharp increase in the curve at low false positive rates; therefore, high-risk patients will be identified with fewer false positives. SHAP attributions are shown for the representative prediction in Figure 2, where the significant contributors are AST, NAFLD, ALT, GGT, and age.



Figs. 1, 2. Left: ROC curve of the proposed model. Right: SHAP attribution for a representative prediction.

The confusion matrix (Table 4) shows most early fibrosis cases classified correctly, though some advanced cases are missed, as is common where feature distributions overlap.

Table 4. Confusion matrix.

	Predicted 0	Predicted 1
Actual 0	102	14
Actual 1	13	18

Threshold analysis (Table 5) exposes the trade-off: at 0.3 recall reaches 0.8889, suiting screening, at the cost of precision; at 0.5 accuracy and F1 are highest; at 0.7 precision and specificity reach 1.0000 but recall falls to 0.4444. Threshold choice should follow clinical priority [19].

Table 5. Effect of the decision threshold on model performance.

Threshold	Accuracy	Precision	AUROC	Recall	F1	Specificity	Sensitivity
0.3	0.7963	0.5581	0.9289	0.8889	0.6857	0.7654	0.8889
0.5	0.8981	0.9000	0.9289	0.6667	0.7660	0.9753	0.6667
0.7	0.8611	1.0000	0.9289	0.4444	0.6154	1.0000	0.4444

5. DISCUSSION

Use of routine biomarkers in non-invasive prediction of fibrosis proved effective. Ensemble methods demonstrated the highest efficiency due to their ability to model interactions among patient's characteristics and use of SHAP allowed every prediction to be explainable. In comparison with FIB-4 the proposed framework achieved higher discrimination (ROC-AUC 0.8529 compared to 0.8493) and accuracy; the difference is slight but significant in the context of mass screening and in accordance with previous researches that have achieved ROC-AUC between 0.82 and 0.89. Prediction based on SHAP scores conforms to the

clinical intuition, since AST/PLT ratio, age, BMI and fasting glucose were found to have the highest impact. The web-based prediction system providing real-time predictions allows using the model for clinical practice [20]. There are two drawbacks: lack of representativeness of the data set in terms of NAFLD patients and oversimplification of the disease through binary classification.

6. CONCLUSION

In this paper, we have provided an interpretable machine learning model to predict non-invasive liver fibrosis based on biomarkers. Ensemble models performed better than traditional statistical models and even slightly better than FIB-4, while SHAP explained the reasons for each prediction. This framework is very useful in developing countries like Southeast Asia, where there is no access to advanced diagnostics tests.

REFERENCES

[1] B. J. Perumpail, et al., 'Clinical epidemiology and disease burden of NAFLD', World J. Gastroenterol., vol. 23, pp. 8263–8276, 2017.

[2] Z. M. Younossi, et al., 'The economic and clinical burden of NAFLD in the United States and Europe', Hepatology, vol. 64, pp. 1577–1586, 2016.

[3] Z. M. Younossi, et al., 'Global epidemiology of NAFLD', Hepatology, vol. 64, pp. 73–84, 2016.

[4] S. A. Polyzos, J. Kountouras, C. S. Mantzoros, 'Obesity and NAFLD: from pathophysiology to therapeutics', Metabolism, vol. 92, pp. 82–97, 2019.

[5] F. Nassir, 'NAFLD: mechanisms, treatments, and biomarkers', Biomolecules, vol. 12, art. 824, 2022.

[6] M. Charlton, 'Non-alcoholic fatty liver disease: current understanding and future impact', Clin. Gastroenterol. Hepatol., vol. 2, pp. 1048–1058, 2004.

[7] S. Y. Park, et al., 'AST/ALT ratio and platelet counts as predictors of fibrosis', Korean J. Gastroenterol., vol. 43, pp. 246–251, 2004.

[8] R. K. Sterling, et al., 'A simple non-invasive index to predict significant fibrosis in HIV/HCV coinfection', Hepatology, vol. 43, pp. 1317–1325, 2006.

[9] A. Robinson, R. J. Wong, 'Noninvasive methods for evaluating hepatic fibrosis in NAFLD', Clin. Liver Dis., vol. 15, no. 4, pp. 157–161, 2020.

[10] A. Esteva, et al., 'A guide to deep learning in healthcare', Nat. Med., vol. 25, no. 1, pp. 24–29, 2019.

[11] M. S. Alam, et al., 'Comparison of FIB-4, APRI and AST/ALT ratio with FibroScan in NAFLD', Egypt. Liver J., vol. 15, art. 38, 2025.

[12] C. Sang, et al., 'Diagnosis of fibrosis using blood markers and logistic regression in NAFLD', Front. Med., vol. 8, art. 637652, 2021.

[13] H. Zamanian, et al., 'AI techniques for NAFLD diagnosis: a systematic review', Comput. Methods Programs Biomed., vol. 244, art. 107932, 2024.

[14] G. Feng, et al., 'Machine learning outperforms fibrosis markers in biopsy-confirmed NAFLD', J. Hepatobiliary Pancreat. Sci., vol. 28, pp. 593–603, 2021.

[15] Z. Zhang, et al., 'XGBoost-based explainable model for predicting significant fibrosis in autoimmune hepatitis', QJM, vol. 119, pp. 34–41, 2026.

[16] T. Chen, C. Guestrin, 'XGBoost: a scalable tree boosting system', Proc. KDD, 2016.

[17] K. Hajian-Tilaki, 'ROC curve analysis for medical diagnostic test evaluation', Caspian J. Intern. Med., vol. 4, no. 2, pp. 627–635, 2013.

[18] S. M. Lundberg, S. I. Lee, 'A unified approach to interpreting model predictions', Proc. NeurIPS, 2017.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Deep Learning-Driven Multi-Omics Integration for Cancer Classification and Biomarker Discovery

Pakhi Sharma[1], Namrata Dhanda[2], Vandana Yadav[3], and Krishna Yadav[4]
[1,4]*Department of CSE, Amity University, Lucknow, U.P., India*
[2]*Department of CSE, SAREC, Mirzapur, U.P., India*
[3]*Department of CSE, Brainware University, Kolkata, W.B., India*
paksh2004@gmail.com | ndhanda510@gmail.com | vandanayadav771@gmail.com | yadav.krishna.tech@gmail.com

Abstract. Breast cancer survival-status classification is difficult because tumour biology varies at several molecular levels at once, which single-omics data rarely captures. This paper integrates copy number variation (CNV), somatic mutation (MU), gene expression (RS) and protein expression (PP) data for survival prediction. Modality-specific autoencoders learn a compact latent representation of each omics layer, and the latent vectors are fused so complementary information is modelled together and used to predict survival status. On a TCGA breast cancer cohort of 705 samples the model reached a test accuracy of 85.1%, an F1-score of 0.222 and a ROC-AUC of 0.582, indicating that latent multi-omics representations carry useful signal while validation on balanced, independent datasets is still needed. Encoder weights also allow candidate biomarkers to be identified.

Keywords. Multi-omics integration; deep learning; cancer classification; autoencoder; biomarker discovery; TCGA; gene expression; copy number variation; proteomics.

1. INTRODUCTION

Breast cancer varies at the genomic, transcriptomic and proteomic levels. High-throughput sequencing has made these changes easier to examine [1], but a single data type gives only a partial view. Combining omics layers allows the disease to be studied from several perspectives and identifies patterns tied to different subtypes and outcomes [2,3]. Because omics datasets carry very large feature counts, machine learning and deep learning are widely applied to cancer classification and survival prediction [2–18]. High dimensionality remains a difficulty, and the models are hard to interpret.

This study integrates CNV, MU, RS and PP data with autoencoders. Because each dataset own structure, a separate autoencoder is trained per omics type, compressing it into a latent representation that retains informative content while suppressing noise [3,4]; the representations are fused and used to predict survival. The contributions are: integration of four omics sources for survival-status classification; modality-specific autoencoders for compact representation learning; a comparison of single-omics, early-fusion and autoencoder-based approaches; and exploration of the learned features for candidate biomarker identification.

2. LITERATURE REVIEW

Azmi et al. [1] compared a stacked denoising autoencoder with a variational autoencoder and assessed feature selection; the variational model reached 91.86% accuracy on lung squamous cell carcinoma. Alharbi et al. [2] applied graph neural networks to biological relationships expressed as graphs, beating conventional methods at higher cost. Subramanian et al. [3] reported that integration generally improves on individual data types, while noting imbalanced data and limited interpretability. Heo et al. [4] found that jointly using genomics, transcriptomics, proteomics and epigenomics improves subtype classification, patient grouping and prognosis, though differences between data types complicate integration. Choi and Chae [5] learned shared patterns across molecular sources for breast cancer, constrained by limited sample numbers.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. DATASET DESCRIPTION

The data come from The Cancer Genome Atlas (TCGA), widely used for its standardised molecular profiles. Mutation data capture genetic alterations, CNV data show genomic gains and losses, gene expression reflects transcription, and proteomic data describe protein-level changes. The cohort has 705 samples with 1,936 features each (Table 3.1); the target is the "vital status" field, encoded as a binary outcome (Table 3.2).

Tables 3.1, 3.2. Omics distribution (left) and survival status distribution (right).

Omics type	Features
Mutation (MU)	249
Copy Number Variation (CNV)	860
Gene expression (RS)	604
Protein-level profile (PP)	223

Survival status	Patients
Survived (0)	611
Not survived (1)	94
Total	705

4. METHODOLOGY

4.1. Preprocessing, Scaling and Splitting

The dataset was organised into a separate matrix per omics type, with survival status stored separately to prevent leakage. Z-score normalisation, giving zero mean and unit standard deviation, was applied so no feature dominates training through its range alone:

$$Z = \frac{X - \mu}{\sigma} \tag{4.1}$$

where μ is the feature mean, σ the standard deviation and Z the standardised value. Scaling was applied only to PP and RS: CNV values are discrete codes (0, 1, −1, 2, −2) whose biological meaning scaling would destroy, and MU values are binary. The data were split 70:15:15 into training, validation and test sets by stratified sampling, preserving the class distribution.

4.2. Model Development

Single-omics baselines were trained first with logistic regression, chosen for its interpretability on high-dimensional data. Early fusion followed, concatenating the four matrices into one representation so a model could learn across layers in parallel; logistic regression and random forest were both applied. Such classifiers may miss non-linear relationships, so a deep fusion model was introduced. An autoencoder's encoder maps input into a smaller latent space and its decoder reconstructs it, so minimising reconstruction error teaches it to keep structure and discard noise. Separate autoencoders were trained on CNV, MU, RS and PP data, their latent features concatenated, and a weight of 1.5 applied to the PP representation before classification (Table 4.1). Features that most influence the prediction, traced through the encoder weights, are treated as candidate biomarkers.

Table 4.1. Autoencoder and training configuration.

Parameter	Configuration	Parameter	Configuration
Encoder	256 neurons, ReLU	Classifier	128 neurons, ReLU
Encoder dropout	0.3	Classifier dropout	0.4
Latent dimension	CNV 64, MU 32, RS 64, PP 32	Output layer	1 neuron, sigmoid
Decoder	256 neurons, ReLU	Optimiser / LR	Adam / 0.001

Parameter	Configuration	Parameter	Configuration
Reconstruction layer	Linear	Batch size / epochs	32 / 50
Fusion dimension	192	Classification loss	Binary cross-entropy
PP latent weight	1.5	Reconstruction loss	MSE (weight 0.05)

5. PERFORMANCE METRICS

Accuracy, F1-score and the area under the receiver operating characteristic curve (ROC-AUC) were used. Accuracy is the proportion of correct predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

where TP, TN, FP and FN are true and false positives and negatives. Accuracy misleads under class imbalance, so the F1-score was recorded too; it combines precision (share of predicted positives that are correct) with recall (share of actual positives recovered):

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5.4}$$

ROC-AUC is the area under the curve of true positive rate against false positive rate; higher values indicate stronger, threshold-independent discrimination.

6. RESULTS AND DISCUSSION

Results are summarised in Table 6.1. The dataset is heavily imbalanced, 611 survived against 94 not-survived, so accuracy alone describes performance poorly; the low F1-scores show how hard the minority class is.

Table 6.1. Comparative performance across modelling approaches.

Model	Accuracy	F1-score	ROC-AUC
CNV only (LR)	0.761	0.000	0.349
MU only (LR)	0.717	0.167	0.559
RS only (LR)	0.761	0.261	0.566
PP only (LR)	0.435	0.435	0.716
Early fusion (LR)	0.704	0.087	0.516
Early fusion (RF)	0.887	0.200	0.628
AE fusion (proposed)	**0.851**	**0.222**	**0.582**

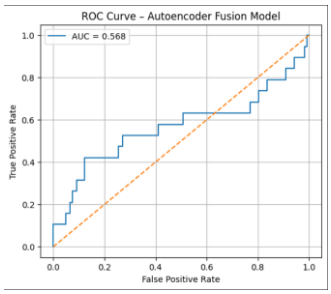


Fig. 6.1. ROC curve for the proposed autoencoder fusion model.

Each omics dataset supplies useful information while fusion gives a broader view, and the latent representations may point to genes and proteins worth following up as biomarkers, subject to independent validation. Limitations: only TCGA data were used; the latent features are hard to interpret; class imbalance affects the F1-score; and the deep model costs more computation.

7. CONCLUSION AND FUTURE WORK

This study built a multi-omics classification framework from copy number variation, somatic mutation, gene and protein expression data in TCGA. The autoencoder fusion model achieved a slightly higher F1-score than the early-fusion random forest, which nonetheless remained ahead on accuracy and ROC-AUC. Future work will test the biomarkers on independent datasets and through pathway analysis, add DNA methylation and miRNA expression, and apply explainable AI to identify which genes and omics layers drive the prediction. Better handling of class imbalance is the most direct route to stronger results.

8. REFERENCES

[1] N. S. Azmi, et al., 'Deep learning for cancer classification using multi-omics feature selection', Prog. Microbes Mol. Biol., vol. 5, no. 1, 2022.

[2] F. Alharbi, A. Vakanski, B. Zhang, M. K. Elbashir, M. Mohammed, 'Multi-omics integration using graph neural networks for cancer classification', IEEE Access, vol. 13, pp. 37724–37736, 2025.

[3] I. Subramanian, et al., 'Multi-omics data integration, interpretation, and its application', Bioinform. Biol. Insights, vol. 14, 2020.

[4] Y. J. Heo, et al., 'Integrative multi-omics in cancer research: networks to clinical subtypes', Mol. Cells, vol. 44, no. 7, pp. 433–443, 2021.

[5] J. M. Choi, H. Chae, 'moBRCA-net: breast cancer subtyping using multi-omics attention networks', BMC Bioinformatics, vol. 24, no. 1, 2023.

[6] Y. Lin, et al., 'Breast cancer subtypes via deep neural networks on multi-omics data', Genes, vol. 11, 2020.

[7] Y. Ektefaie, W. Yuan, D. A. Dillon, et al., 'Integrative multiomics-histopathology analysis for breast cancer classification', npj Breast Cancer, vol. 7, 2021.

[8] A. F. Abbasi, et al., 'Multi-omics framework for cancer molecular subtyping', Sci. Rep., vol. 15, 2025.

[9] A. M. Hassan, S. M. Naeem, M. A. A. Eldosoky, et al., 'Multi-omics machine learning for breast cancer subtyping', Arab. J. Sci. Eng., vol. 50, 2025.

[10] D. Leng, L. Zheng, Y. Wen, et al., 'Benchmark of deep learning multi-omics fusion methods for cancer', Genome Biol., vol. 23, 2022.

[11] J. Zhang, Y. Che, R. Liu, Z. Wang, W. Liu, 'Deep learning-driven multi-omics analysis for cancer diagnostics', Brief. Bioinform., vol. 26, no. 4, 2025.

[12] D. Wang, J. Gu, 'Integrative clustering of multi-omics data for cancer classification', Quant. Biol., vol. 4, 2016.

[13] J. S. Wekesa, M. Kimwele, 'Multi-omics deep learning for disease diagnosis', Front. Genet., vol. 14, 2023.

[14] B. Ma, et al., 'Cancer classification using XGBoost and multi-omics data', Comput. Biol. Med., vol. 121, 2020.

[15] M. Tao, T. Song, W. Du, S. Han, C. Zuo, Y. Li, Y. Wang, Z. Yang, 'Breast cancer subtypes using multiple kernel learning on omics data', Genes, vol. 10, 2019.

[16] L. Tong, J. Mitchel, K. Chatlin, et al., 'Feature-level integration of multi-omics data for breast cancer survival', BMC Med. Inform. Decis. Mak., vol. 20, 2020.

[17] Z. Huang, X. Zhan, S. Xiang, et al., 'SALMON: survival analysis learning with multi-omics neural networks', Front. Genet., vol. 10, 2019.

[18] S. Yang, et al., 'Machine learning for subtyping triple-negative breast cancer', J. Cancer, vol. 15, 2024.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Comprehensive Deep Learning Approach for Heart Failure Risk Assessment and Classification

Zainab Hassan[1], Namrata Dhanda[1], Vandana Yadav[2] and Neha Chauhan[1]

[1]*Department of Computer Science & Engineering, Amity University, Lucknow, Uttar Pradesh, India*
[2]*Department of Computer Science & Engineering, Brainware University, Kolkata, West Bengal, India*
zainabhassan95561@gmail.com, ndhanda@lko.amity.edu, vandanayadav771@gmail.com, Neha.chauhan8924@gmail.com

Abstract.
Heart failure is a chronic and progressive cardiovascular condition that contributes substantially to morbidity and mortality worldwide. Early identification of the disease remains difficult because its progression varies among patients and may be associated with a wide range of clinical characteristics. At the same time, the increasing availability of healthcare data has created new opportunities for applying artificial intelligence to medical prediction and classification. In particular, deep learning methods are being investigated for identifying patients who may be at risk of heart failure and for improving the classification of disease-related outcomes.

The present study develops a deep learning-based framework for analysing heart failure using patient clinical records. The approach focuses on identifying patterns within patient data that may be associated with adverse heart failure outcomes. Artificial Neural Networks (ANNs) and other deep learning approaches are considered for their ability to model relationships among multiple clinical variables. The study also discusses commonly used machine learning techniques, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Gradient Boosting, to provide a broader view of approaches used for heart failure classification.

Model performance can be assessed using measures such as accuracy, precision, recall, F1-score, and AUC-ROC. These metrics provide different perspectives on classification performance and are particularly useful when evaluating medical prediction models. The analysis also emphasizes the importance of data preprocessing and feature preparation before model development, since the quality and consistency of the input data can influence the resulting predictions.

Conventional approaches to heart failure assessment generally involve clinical symptoms, laboratory investigations, and medical imaging. Although these methods remain important in clinical practice, analysing large amounts of patient information can be challenging. Machine learning and deep learning provide computational approaches for examining larger datasets and identifying relationships among multiple variables. Therefore, these techniques may offer useful support for early risk assessment and classification of heart failure, while further validation with larger and more representative clinical datasets remains necessary.

Keywords. Heart disease, deep learning, artificial neural networks, risk assessment, disease classification, hyperparameter tuning.

1. INTRODUCTION

Heart Failure (HF) is a serious and progressive disease characterized by the reduced ability of the heart to pumps blood that meets the body's requirements and is related to high morbidity rates. There is an immediate need in the identification of patients with a risk of developing HF or predicting their further course of disease.

At the moment, the task of diagnosing HF includes both clinical examinations, various laboratory tests, and imaging studies. Although these conventional diagnostic methods are effective, they are often resource-intensive and lengthy procedures. More importantly, there are a number of patterns in patients' data that cannot be detected using traditional methods.

At the same time, in combination with the growth of the number of patients with chronic diseases, the introduction of electronic health records and the collection of medical databases has enabled the use of computational methods for assisting medical diagnoses.

Machine learning algorithms have demonstrated great potential for analyzing large sets of patient data and detecting patterns useful in clinical practice. However, most conventional machine learning techniques rely mainly on manually engineered features and have limited abilities in learning complex associations between different attributes.

Deep learning models, on the other hand, enable modeling of complex associations and can help detect patterns that conventional methods fail to detect. Deep learning technology has been introduced as a powerful machine learning approach capable of achieving state-of-the-art results in many challenging pattern recognition tasks. In particular, artificial neural networks showed promising results in clinical applications. For example, they can be used to study the relationship between multiple patient characteristics, such as age, blood pressure, cholesterol levels, heart rate, and other indicators to find patterns associated with the presence of HF. Moreover, a variety of additional features can be added to the model to account for other factors that may affect the development of the disease.

One of the main advantages of using deep learning models is the ability to work with high-dimensional data of different types. In particular, a large amount of data from various sources, including structured data from clinical questionnaires, laboratory tests, physiological measurements, and time series of patient vitals, can be used. The complexity of the interactions between different factors is well known and is one of the challenges for developing and testing appropriate mathematical models. At the same time, artificial neural networks are universal approximators capable of learning complex associations between different attributes. This property is especially useful when working with clinical data since most diseases, including HF, are caused by a combination of different factors. The ability to learn complex interactions between different characteristics makes deep learning models attractive for solving clinical problems.

Another advantage of using deep learning approaches in biomedical research is the ability to simultaneously process multiple types of data. Clinical studies include a vast amount of different types of information, from structured data from clinical questionnaires to time series of patient vital signs and laboratory test results and various imaging data. Each of these types of information usually requires special processing before use in conventional modeling approaches. At the same time, the use of artificial neural networks makes it possible to simultaneously process different types of information using appropriate network architectures. For example, such an approach can be used to process structured data from clinical questionnaires and laboratory tests, time series of patient vitals, and echocardiographic images. Such an integrated approach allows for creating more efficient models since different types of information can be combined in an optimal way to maximize the information content.

Thus, the current project is aimed at developing a systematic deep learning model capable of estimating and classifying the risks of HF. Of primary importance is the selection of a neural network architecture capable of efficiently analyzing the input data while providing sufficient accuracy in predicting the probability of HF. To do this, it is necessary to prepare a set of input data, preprocess them, configure a simple neural network, train it, and evaluate its performance. Applying deep learning approaches to solve clinical problems will allow expanding the capabilities of intelligent medical information systems, improving the accuracy of medical assessments, reducing the time spent on them, and thereby improving the quality of care for patients with chronic diseases such as HF. In particular, this project will demonstrate how deep learning models can be used to predict the development of HF based on patient data.

2. LITERATURE WORK

In recent years, machine learning and deep learning have attracted the attention of researchers who study the prediction of heart failure because of the data on patients' health conditions. **Qadeer et al. (2025)** compared several machine learning and deep learning models for the task of heart failure prediction and evaluated their performance using clinical data [1]. Their research shows that computational models can help clinicians with analyzing the collected health data, especially when multiple patient-specific attributes need to be taken into consideration. **Norouzi et al. (2024)** proposed a deep learning model that tackles the survival prediction problem in patients with heart failure by addressing both the notion of frailty and competing risks [2]. The researchers demonstrated the complexity of making clinical predictions about the survival of heart failure patients. Additionally, **Scientific Research Publishing (2023)** evaluated a combination of machine learning and deep learning methods for the classification of heart failure [3]. **Choi**

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

et al. (2017) developed a deep learning model known as the RETAIN, which utilizes a reverse-time attention mechanism for the purpose of creating predictions based on the patient's healthcare data [4]. One of the main advantages of this model is the fact that it allows for distinguishing the most relevant healthcare records, thus enabling the interpretation of the final results. **Rajkomar et al. (2018)** were the first who managed to explore the potential of using deep learning and electronic health records for making various predictions regarding patients' conditions [5]. **Yao et al. (2019)** employed a knowledge-guided convolutional neural network to solve the task of clinical text classification, which helped them address the particular challenge of making predictions based on textual health data [6]. The study illustrates that the analysis of clinical notes can be beneficial because the information that is contained in them is often essential and cannot be found elsewhere. **Krittanawong et al. (2020)** highlight the importance of using machine learning technologies in cardiovascular medicine, as well as discuss their potential benefit for making survival predictions and providing the best patient outcomes [7].

Alaa and van der Schaar (2018) proposed a deep learning-based approach called DeepHit that can be used to predict the survival time of patients with competing risks [8]. **Esteva et al. (2019)** discussed the potential of deep learning algorithms in different areas of medical practice, as well as summarized the ways in which various deep learning-based solutions were used for different prediction and diagnostic tasks [9]. Overall, the authors showed that deep learning technologies can be used in different medical fields despite the variability of input information, task goals, and the final objectives. The use of sequential information about the patient's health condition has attracted the attention of several researchers. For instance, **Lipton et al. (2016)** evaluated recurrent neural networks' ability to learn sequentially structured clinical records and addressed the problem of missing data [10]. **Johnson et al. (2018)** highlighted the role of artificial intelligence and machine learning algorithms in modern cardiology and discussed their potential for clinical use [11]. **Attia et al. (2019)** addressed the possibility of using an artificial intelligence algorithm to analyze electrocardiograms and detect patients who are at risk of developing heart failure [12]. **Cho et al. (2014)** designed an encoder-decoder model that is based on recurrent neural networks and can be used for various sequential tasks [13]. **Miotto et al. (2016)** introduced the Deep Patient framework, which was created using an unsupervised deep learning algorithm that managed to learn the representations of electronic health records and perform multiple prediction tasks [14]. Finally, **Zheng et al. (2021)** reviewed the current research papers and described the deep learning methods used for predicting heart failure readmission [15].

3. USE OF ALGORITHMS

3.1. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) represent one of the fundamental approaches to deep learning that could be implemented for medical information analysis. The network comprises multiple processing layers of interconnected neurons, where the layers typically comprise an input, a hidden, and an output layer. Each neuron receives input from other neurons through weighted connections and uses an activation function to produce an output signal.

For the prediction of heart failure, ANN could be applied to the structured data, which include demographic data, blood pressure, serum creatinine, ejection fraction, and other laboratory metrics. Using multiple variables, it would be possible to explore associations between different risk factors in a multivariate model. Additionally, the network would allow considering numerous patient characteristics related to the condition, thus providing a sufficient exploration of clinical information.

Using the backpropagation procedure, it would be possible to train the network and optimize the weights' distribution in order to minimize the difference between the true and predicted outcome. In this network, the Rectified Linear Units (ReLU) would be used as an activation function for the hidden layer, and a Sigmoid function could be applied to the output layer in order to predict the class probabilities for the binary classification. Overall, ANNs' implementation in the heart failure prediction task would enable building a relatively simple model that could serve as the basis for the solution. Still, it should be noted that the approach's success depends on the data amount and quality, network architecture, and other parameters that should be optimized during the training process.

3.2. Convolutional Neural Network

The CNNs are created for the analysis of the data that have some relationships either in space or time. Hence, CNNs can be helpful in the medical applications of images and signals, particularly, ECG signals. In general, a CNN contains the convolutional layers where different filters are applied to the input data and the patterns and features are identified. These layers can be followed by the pooling operations that allow reducing the dimensions of the data. The CNNs can help in analysing the ECG signals and echocardiographic data in relation to the heart failure research. The CNNs can recognize the patterns that are contained in the forms of the medical data. This recognition cannot be performed using other approaches. In addition, another great benefit of the CNNs is that they are able to learn some features directly from the input data.

Nevertheless, the CNNs have some drawbacks. The training of the CNNs requires a sufficient number of data and can consume much computing resources. The performance of CNNs is influenced by the quality of the input data, network architecture, and training. Therefore, the proper preprocessing and configuration of the CNNs are essential in relation to the heart failure medical data.

3.3. Feed Forward Deep Neural Network

This network is an extended version of the basic Artificial Neural Network (ANN). It has multiple (usually more than 2) hidden layers allowing the model to extract more complex patterns from the input data. Information in the network moves only in one direction from inputs to outputs without backtracking.

Deep Neural Networks are particularly effective in cases where the data consists of many features, such as in the case of clinical data in heart failure patients.

Such networks can easily learn patterns related to different factors that contribute to the condition and better understand the relationships between various inputs. For instance, a DNN can learn to process the information about hypertension, diabetes, and the patient's age and determine how each of these factors taken together affects the likelihood of developing heart failure. At the same time, the complexity of such networks is their disadvantage as too great complexity in relation to the available data will result in overfitting. Thus, proper tuning of the layers and neurons number, learning rate, and other hyperparameters, and the application of dropout regularization are crucial for training such models.

3.4. Recurrent Neural Network

The described below artificial intelligence structures are primarily applied to temporal data as they allow performing the task better than the standard feed-forward neural networks. The main difference consists in having recurrent connections that enable them to learn and retain the hidden state from previous computations. It becomes possible due to the fact that in case of sequential data, the information from earlier steps is required to compute the following ones.

For example, such neural networks can be utilized in medical studies dealing with patients treatment. Specifically, recurrent networks can process the set of patients visits data with particular tests and symptoms as inputs. Moreover, RNNs can operate on the long-term temporal data, which is essential for instance for cardiovascular diseases progress analysis. Nonetheless, vanilla RNNs suffer from a vanishing gradient problem that prevents learning long-term dependencies. For this reason, more advanced Long Short-Term Memory networks and Gated Recurrent Units were developed.

3.5. Long Short Term Memory

LSTM networks, which make use of memory cells and control gates to manage information flow, enable them to learn long-term dependencies better than other neural networks. Such networks have been created to address the shortcomings of standard recurrent neural networks. When it comes to heart failure prediction, LSTM networks can study the progression of the disease over extended periods of time. In other words, the patterns discovered by these networks could be used to extrapolate changes in blood pressure or heart rate that could lead to a heart failure diagnosis. Apart from that,

compared to standard models, LSTM networks tend to have better performance when it comes to prediction tasks. At the same time, their design is more complicated, which means that the implementation and training process could be more challenging.

4. Applications of Deep Learning in Heart Failure Risk Assessment

1. **Initial Risk Assessment and Preventive Healthcare-** Deep learning algorithms hold great promise for predicting patients likely to suffer from heart failure before the onset of clinical signs. They analyze a number of clinical parameters collectively including laboratory findings, demographics, comorbidities, vital signs as well as patient history. For this reason, they are able to pick up on subtle relationships between different variables that would normally be overlooked by traditional approaches. The value of early risk prediction lies in the fact that it helps physicians initiate preventive measures at the appropriate time. Such preventive measures could range from modifying drug regimens, adopting a healthier lifestyle and follow-ups among others.
2. **Clinical Decision Assistance Systems-** Incorporating deep learning algorithms in hospital management systems will enable practical implementation of this technology in the sphere of medicine. For instance, such algorithms will be able to evaluate patient data and estimate the score or classify the data, thus giving an insight into the difficulty of the case and possible complications. In emergency cases, it may be helpful to predict the need in assistance. Furthermore, automation of such a system will reduce the burden on clinicians and allow them to dedicate their time to other important issues. It should be noted that such algorithms are meant to support medical staff in their work and not to substitute them.
3. **Customized Treatment Strategy-** Heart failure differs from patient to patient since it is influenced by a number of different aspects, including age, genetics, other diseases, and history of treatment among others. With the help of deep learning algorithms, it would be possible to discover the patterns in patient's medical history from which it would be possible to find out what kind of therapy is more suitable for him or her. At the same time, it would be also possible to choose a therapy for a patient that differs from the therapy chosen for everybody else. In this regard, it can be stated that these technologies actually contribute to the personalization of treatment, which is a core of precision medicine.
4. **Ongoing Surveillance and Remote Healthcare Solutions-** Technological improvements in wearable devices plus remote monitoring systems have made it possible to collect continuous physiological data from patients even outside hospitals. Using this real-time information, deep learning algorithms can help clinicians identify patterns that may show worsening of a patient's condition. Changes in heart rate, blood pressure, along with oxygen saturation can be monitored continuously to detect early warning signs. Besides this, such systems may help doctors respond faster before serious complications develop. Continuous monitoring also gives a clearer picture of patient health over time instead of depending only on occasional hospital visits. This can improve patient care as well as support better management of heart-related conditions.
5. **Prediction of Readmission and Mortality-** Readmission rates along with death rates still remain major problems for patients suffering from heart failure. Deep learning algorithms can help in predicting the chances of negative outcomes by studying patterns found in discharge summaries, medications, laboratory reports, plus records of previous hospital admissions. These models may identify high-risk patients earlier compared to traditional approaches. Besides this, proper prediction allows healthcare organizations to plan targeted care programs for patients who need extra attention. Such support can include closer monitoring, medication management, as well as regular follow-up care. In many cases, this may help improve survival rates while also making healthcare delivery more effective and organized.

5. Methodology

1. To begin with, the proposed method is based on deep learning algorithms that can evaluate and classify the risks of heart failure. In order to prepare the ground for the study, the data set consisting of the records of 300 patients was used. The data set includes such independent variables as age, ejection fraction, serum creatinine, serum sodium, platelet count, and time of the follow-up that are associated with the main diagnosis of heart failure. In this study, the dependent variable was referred to as "DEATH_EVENT," which implies a binary outcome of either being dead or alive ("1" refers to death and "0" refers to being alive).
2. Exploratory data analysis then took place. I have plotted a histogram of each of the clinical parameter. The correlation heatmap was plotted to analyze the feature interaction and whether there exists a correlation level between any two features concerning the target variable.

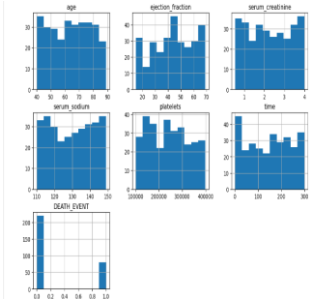


Figure 5.1. Histogram of heart failure features

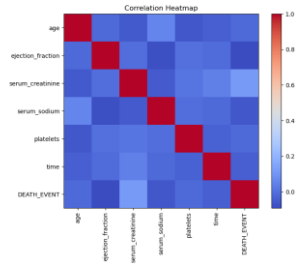


Figure 5.2. Correlation heatmap

3. Once the visualization process was complete, I moved on to the data preprocessing phase. At this stage, I isolated the input variable (X) from the output variable (y). Additionally, an 80/20 split was carried out on the dataset. This implies that 80% of the data was used for training while 20% of the data was used to test the performance of the model. Other than conducting an 80/20 data split, normalization of the data was done by standardizing the range of the variable. This step is important when using a neural network model since it makes the training faster and enhances the model's performance.
4. The classification algorithm was implemented by using Multi Layer Perceptron (MLP) from the Scikit-learn machine learning library using MLP Classifier. The network consists of two hidden layers having 16 neurons in the first layer and 8 neurons in the second one. ReLU was chosen as the activation function for the network whereas the Adam optimizer was used for training the network.

5. Lastly, the model was evaluated using accuracy score along with a confusion matrix. The accuracy score showed the overall prediction performance, while the confusion matrix provided values like true positive, true negative, false positive, plus false negative. These evaluation methods helped in checking the effectiveness of the deep learning classification model step by step.

6. Testing and Analysis

The performance of the proposed model was evaluated using an 80/20 split between training plus testing data. Around 80% of the dataset was used for training the model, while the remaining 20% was reserved for testing so that the model could be checked on unseen data samples. Before starting the training process, all variables were standardized to remove differences caused by unequal data scaling. This step helped improve learning efficiency as well as model stability. The proposed neural network architecture was based on a Multi-layer Perceptron containing two hidden layers. ReLU activation functions were applied in the network, whereas the Adam optimization method was used during training to improve the updating of weights inside the model.

7. Result and Discussion

The designed model for heart failure risk classification was analyzed in terms of a well-structured train-test scheme. Training the Multi-Layer Perceptron on the chosen clinical characteristics, the performance metrics on the test sample were estimated. In other words, the level of precision proves that the ability to classify most of the cases properly was demonstrated by the model, implying the existence of the relevant pattern in the training set.

Nevertheless, in order to figure out how the model works in detail, it is crucial to analyze the confusion matrix. One should point out that regardless of several mistakes, the model is able to quite efficiently split patients into two categories depending on whether they have the risk of heart failure or not. Nevertheless, the mentioned mistakes are quite understandable due to the imbalanced data and limited dataset. At the same time, it is worth noting that the results can be supplemented with graphical analysis using histograms and the correlation heatmap. This analysis allows better visualization of the problem since histograms represent the distribution of individual characteristics, while the heatmap reveals the dependencies between variables.

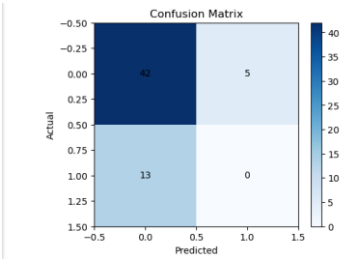


Figure 7.1. Confusion matrix of the classification model

8. Conclusion and Future Scope

In conclusion, this study analyzed a deep learning-based algorithm for the evaluation of risk of heart failure by choosing appropriate patient attributes for classification. The artificial neural network has been developed and tested based on the knowledge of the correlation between the patient parameters and his/her health status. It was determined that the proposed algorithm has proven itself quite effective in terms of providing reliable predictions, suggesting its ability to detect important dependencies from the data.

Preprocessing has been used in order to ensure consistency of all patient features in the dataset. Feature scaling, which has been applied to data prior to analysis, allowed for stabilizing the model training process. An additional analysis of the model performance was conducted via metrics such as the calculation of accuracy and confusion matrices, revealing certain inconsistencies, including some misclassifications of patients' status.

The next step to take may include the consideration of other, more complex model architectures and the incorporation of more patient data for predicting better. One may also consider the possibility of combining different medical data sources and building predictive systems that work in real-time.

9. References

[1] Qadeer, M., Ayaz, R., Thohir, M. I., 'Heart Failure Prediction Through a Comparative Study of Machine Learning and Deep Learning Models', Engineering Proceedings, vol. 107, no. 1, MDPI, 2025.
[2] Norouzi, S., Khormaei, H., Jafarabadi, M. A., Hajizadeh, E., Naderi, N., 'Development of a Deep Learning Model for Survival Prediction in Heart Failure Using Frailty and Competing Risks', Scientific Reports, Nature, 2024.
[3] Scientific Research Publishing, 'A Novel Approach to Heart Failure Prediction and Classification Through Machine Learning and Deep Learning', 2023.
[4] Choi, E., Bahadori, M. T., Sun, J., Kulas, J., Schuetz, A., Stewart, W. F., 'RETAIN: Interpretable Predictive Model for Healthcare Using Reverse Time Attention Mechanism', Advances in Neural Information Processing Systems, 2017.
[5] Rajkomar, A., Oren, E., Chen, K., et al., 'Scalable and Accurate Deep Learning with Electronic Health Records', npj Digital Medicine, 2018.
[6] Yao, L., Mao, C., Luo, Y., 'Clinical Text Classification with Rule-Based Features and Knowledge-Guided Convolutional Neural Networks', Journal of Biomedical Informatics, 2019.
[7] Krittanawong, C., Virk, H. U. H., Bangalore, S., et al., 'Machine Learning Prediction in Cardiovascular Medicine: Implications for Heart Failure', Journal of the American College of Cardiology, 2020.
[8] Alaa, A. M., van der Schaar, M., 'DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks', AAAI Conference on Artificial Intelligence, 2018.
[9] Esteva, A., Robicquet, A., Ramsundar, B., et al., 'A Guide to Deep Learning in Healthcare', Nature Medicine, 2019.
[10] Lipton, Z. C., Kale, D. C., Wetzel, R., 'Directly Modeling Missing Data in Sequences with RNNs: Application to Predicting Mortality in Intensive Care Units', Machine Learning for Healthcare Conference, 2016.
[11] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., et al., 'Artificial Intelligence in Cardiology', Journal of the American College of Cardiology, 2018.
[12] Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., et al., 'An Artificial Intelligence-Enabled ECG Algorithm for the Identification of Patients with Heart Failure', Nature Medicine, 2019.
[13] Cho, K., van Merriënboer, B., Gulcehre, C., et al., 'Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation', arXiv preprint, 2014.
[14] Miotto, R., Li, L., Kidd, B. A., Dudley, J. T., 'Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records', Scientific Reports, 2016.
[15] Zheng, Y., Peng, H., Zhang, Y., 'Deep Learning Approaches for Predicting Heart Failure Readmission: A Systematic Review', Frontiers in Cardiovascular Medicine, 2021.

Heart Disease Prediction in Explainable AI-Based Clinical Decision Support System based on XGBoost and SHAP

Ravi Ranjan Singh[1], Jatin Kumar[2], Bijoyini Bagchi[2],
Neha Gupta[2], Aditya Singh[2]
Computer Science and Engineering, Institute of Engineering and Management, Kolkata

Abstract: Predicting heart disease using clinical data can provide valuable information for early risk identification, and the difficulty of many machine learning models to interpret is a challenge in healthcare applications. This research proposes an interpretable machine learning model to compare Logistic Regression, Decision Tree, Random Forest, and XGBoost models for predicting heart disease. The clinical data are preprocessed and then split into a training set and a test set, then the model is trained, and finally compared with the accuracy, precision, recall and F1 score. The optimized XGBoost model outperformed the others, making the accuracy (90.76). Subsequently, the clinical features that lead to the prediction of the model were identified using the SHapley Additive exPlanations (SHAP) method. The design proposed here ensures predictive performance and allows for interpretability of the results and it can be used as a potential basis for clinical decision-support applications.

Keywords: Heart Disease Prediction , Machine Learning , Explainable AI , XGBoost , SHAP , Clinical Decision Support.

1 Introduction

Heart disease is still a significant health problem and early risk detection is crucial for early intervention. The clinical data have been extensively used for machine learning techniques like Logistic Regression, Decision Tree, Random Forest and XGBoost. But complex models can be uninterpretable, which means that they cannot be accepted in healthcare. To overcome this limitation, approaches such as LIME or SHAP were created which are able to explain the impact of individual features. The goal of this study is to optimize the performance of the classifiers while maintaining interpretability while accurately predicting the presence of heart disease, this study evaluates four classifiers and combines optimized XGBoost with SHAP.

2 Literature Review

The field of machine learning, clinical data has been employed in numerous applications, such as predicting heart diseases. Detrano et al. created a diagnostic model based on probabilities that contributed to the creation of the well known Heart Disease database of the UCI [1,9]. In classical supervised learning techniques, Kotsiantis studied them and explained their appropriateness for the medical classification problem as well as the fact that they are unable to handle complex nonlinear relationships [6]. Chen and Guestrin proposed an efficient gradient-boosting algorithm, XGBoost, which is good at prediction on structured data [3]. Advanced machine learning models, however, are still a big challenge in healthcare, in being interpretable. Ribeiro et al. [5] proposed the concept of LIME for explaining individual predictions and Lundberg and Lee [4] proposed the concept of SHAP to provide the contribution of the features to the output of the model. However, there is a challenge for research in the development of good predictive models that are clinically relevant. To tackle this, this paper tests a variety of classifiers and uses SHAP for the optimized XGBoost.

3 Methodology

3.1 System Overview

The proposed pipeline comprises six steps: Data collection, Data Preprocessing and Feature Transformation, split- ting the data into (80 percentage) train and (20 percentage) test data, training the model, evaluating the model and explainability analysis using SHAP.

3.2 Dataset and Preprocessing

Other laboratory values and some of the clinical features from each patient are available in these data, for example, whether the person has chest pain, their blood pressure when they're not working out, whether they've got cholesterol, and more. Duplicate records were removed, and only the relevant features were retained and StandardScaler then applied. The dataset was then divided in to two sets: 80% for training and 20% as a test set.

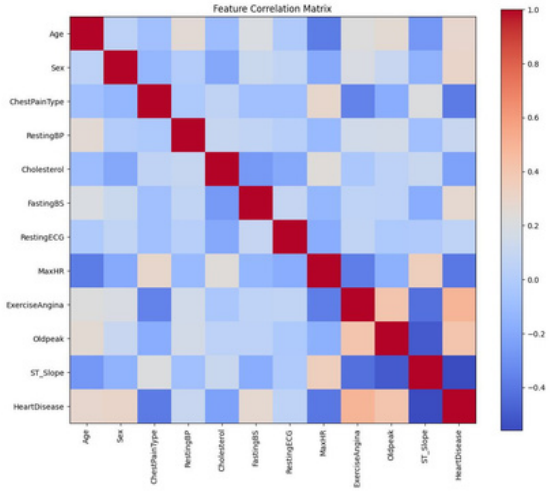


Figure 1: Feature correlation matrix, showing how clinical attributes relate to one another and to the heart-disease target.

3.3 Models and Explainability

The classifiers that were trained are: Logistic Regression, Decision Tree, XGBoost and Random Forest. Hyperparameter optimized XGBoost was used for the main model. All the features were ranked based on their importance using SHAP and individual predictions and overall feature contributions were explained.

3.4 Evaluation Metrics

Accuracy, precision, recall and F1 score were used to assess the performance of the model. Confusion matrix was also worked out for analysing true and false classifications.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \qquad \text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \qquad F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4 Experimental Setup

The experiments have been carried out in python, using Scikit learn and XG boost, in google Colab. Preprocessed data then was split into input features and the target (heart-disease) class, which is binary. RandomizedSearchCV for XGBoost optimization and comparing the classification performances based on the accuracy, precision, recall, F1-score with other classifiers. In addition, a confusion matrix and SHAP feature importance plot was created.

5 Results and Discussion

5.1 ModelComparison

The accuracy, precision, recall and F1 score is demonstrated for the four classifiers in Table 1 below. The XGBoost model was the one that is most optimized and it was seen to perform best with accuracy of (90.76) while the RF model was the second best model. The results showed that the tree-based ensemble models performed better than Logistic Regression and Decision Tree on the given clinical data.

Table 1: Performance comparison of the machine learning models (%).

Model	Accuracy	Precision	Recall	F1Score
LogisticRegression	86.96	83.00	87.00	85.00
DecisionTree	78.80	81.00	84.00	82.00
RandomForest	89.67	89.52	92.16	90.82
XGBoost	90.76	91.26	92.16	91.71

5.2 Confusion Matrix and SHAP Analysis

The confusion matrix for the optimized XGBoost model shows that in most cases, the model had accurately determined if a row had data for a disease or not. SHAP was used to explore how individual clinical metrics relate to the model's predictions. For the analysis of individual clinical features in the model output, SHAP analysis was used. The analysis provides an understandable view of what conditions predict heart disease, and can be combined with the quantitative performance metrics.

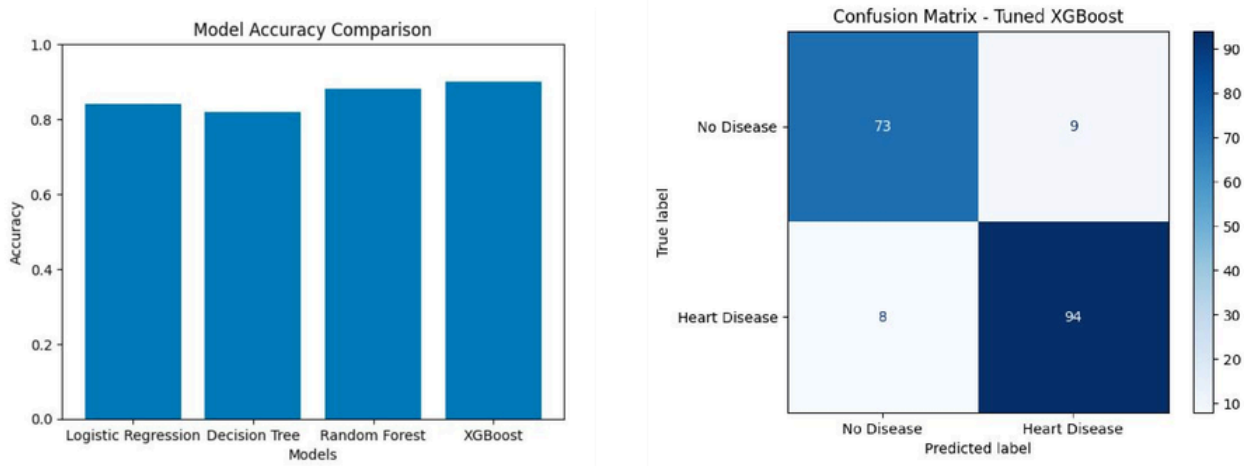


Figure 2: Accuracy comparison across the four classifiers (left) and confusion matrix of the optimized XGBoost classifier (right).

5.3 Discussion

Overall, the results show that the optimized XGBoost classifier is a well-performing machine learning model that enables the effective prediction of the issue, while the SHAP model shows the improved interpretability of the model's predictions. The two are complementary in their development of clinically useful machine learning systems for clinical decision support in healthcare.

6 Conclusion

In this research, machine learning methods were applied for predicting heart disease using clinical parameters. The numeric order of classifiers based on their overall performance was optimized- XGBoost classifiers (90.76%),Random Forest Classifier (89.67%), Logistic Regression Classifier (86.96%) and as the least best was the Decision Tree Classifier with (78.80%). Interpretability of the model was also shown with SHAP giving interpretable view, showing which feature of the model was responsible for each prediction. The results show how well the proposed approach of predictive machine learning and explainable AI can help improve the development of more explainable and reliable clinical decision-support systems.

7 Future Work

Future research will concentrate on enhancing robustness and generalization with larger and more varied clinical datasets that will be used to validate the framework. Further study and comparison might be conducted using other machine learning methods and deep learning methods. Some other ways of explaining could be explored to provide similar explanations. The system might be further extended in the future to a real-time clinical decision-support system to help health care professionals identify risk at an early stage.

References

[1] Detrano, R., et al.: International application of a new probability algorithm for the diagnosis of coronary artery disease. Am. J. Cardiology 64(5), 304–310 (1989)

[2] Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)

[3] Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. Proc. 22nd ACM SIGKDD, 785–794 (2016)

[4] Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. NeurIPS, 4765–4774 (2017)

[5] Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?" Proc. 22nd ACM SIGKDD, 1135–1144 (2016)

[6] Kotsiantis, S.B.: Supervised machine learning: A review. Informatica 31(3), 249–268 (2007)

[7] Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)

[8] Quinlan, J.R.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986)

[9] Dua, D., Graff, C.: UCI Machine Learning Repository. Univ. of California, Irvine (2019)

[10] Lundberg, S.M., Erion, G., Lee, S.I.: Consistent individualized feature attribution for tree ensembles. arXiv:1802.03888 (2018)

[11] Ahmad, M.A., Eckert, C., Teredesai, A.: Interpretable machine learning in healthcare. Proc. 2018 ACM BCB, 559–560 (2018)

[12] Deo, R.C.: Machine learning in medicine. Circulation 132(20), 1920–1930 (2015)

[13] Rajkomar, A., Dean, J., Kohane, I.: Machine learning in medicine. N. Engl. J. Med. 380(14), 1347–1358 (2019)

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Graph-Based Entity Resolution for Real-Time Financial Sanctions Screening: Leveraging Knowledge Graphs and GNN-Enhanced Fuzzy Matching at Enterprise Scale

Suman Reddy Gaddam

Indian Institute of Science & Technology (IIST), India, sumanreddyg05@gmail.com

Abstract.

The problem of financial sanctions screening is to identify entities in financial records, which is typically a huge collection of financial records, from a wide variety of types with inconsistent identity names and confusing textual attributes. The existing rule based and string-matching methods are inadequate to recognize structural relationships between entities, which leads to high false positive rate and low efficiency in operation. The objective of this paper is to propose a framework for real-time financial sanctions screening, called Hybrid Graph-based Fuzzy Entity Resolution (HG-FER). The framework is based on knowledge graph modeling, graph neural network (GNN)-based contextual representation learning, and fuzzy similarity using adaptive fuzzy similarity in a single entity resolution pipeline. Complex relationships are captured in knowledge graphs, their structure is learned by GNNs, and textual uncertainty is manageable with fuzzy matching. Hybrid similarity mechanism: Both relational and attribute level information are used for better entity matching accuracy. Experimental results show that the proposed framework obtains a precision of 0.92, recall of 0.89 and F1-score of 0.90, and has 10 false positives which is better than conventional approaches. The findings show that HG-FER is a scalable, reliable, and effective solution for the enterprise level financial compliance and near real-time sanctions screening.

Keywords. Entity Resolution, Knowledge Graph, Graph Neural Networks, Fuzzy Matching, Financial Sanctions Screening, Fraud Detection.

1. INTRODUCTION

The nature of financial sanctions and the requirement to comply with them has become increasingly complex as financial markets and cross-border transactions have developed and become more digitized. Large data sets of customer and transaction records need to be continuously processed and compared against dynamic sanctions lists, even when the data has different naming conventions, attributes and when some data is missing or noisy. Presently, the conventional sanctions screening systems are fundamentally based on rule-based filtering and string-matching methods that tend to miss out on semantic relationships between entities and thus have high false positive and false negative rates. Such restrictions can drive up operating expenses, compliance risks, and make compliance systems less reliable.

Knowledge graphs and graph neural networks (GNNs) are introduced to represent complex relationships between heterogeneous financial entities and learn the context information

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

from graph structures, which have significantly advanced the representation of entities. But most current approaches are based on one aspect of relational learning—graph-based relational learning—or one aspect of fuzzy attribute matching and don't allow for the accurate disambiguation of ambiguous entities in real-time enterprise settings.

To overcome these shortcomings, the authors introduce in this paper a new framework called Hybrid Graph-based Fuzzy Entity Resolution (HG-FER) for financial sanctions screening in real time. In contrast to traditional approaches that mainly depend on graph-based relational learning or fuzzy attribute matching, HG-FER introduces an adaptive fuzzy similarity and a unified entity resolution pipeline with knowledge graph modeling and GNN-based contextual representation learning. This integrated design not only incorporates structural relationships but also the uncertainty at the attribute level, leading to better matching accuracy, fewer false matches, and enterprise compliance scaling.

This work has the following key contributions:

- The architecture of a knowledge graph-based system to model heterogeneous financial entities and their relations;
- The incorporation of the use of graph neural networks to learn and represent complex relational dependencies and improve the prediction of entity similarity;
- The integration of the fuzzy matching methods to efficiently deal with ambiguous and inconsistent identity attributes and;
- The creation of a scalable architecture to be applied in real-time sanctions screening in enterprise environments;

As far as we know, this work is one of the first to combine fuzzy similarity with graph neural networks in a single framework in terms of real-time financial sanctions screening.

2. RELATED WORK

As financial sanctions screening becomes more complex, intelligent, data-driven approaches to entity resolution and fraud detection have come into use. Knowledge graphs have become a useful method of modelling a combination of financial entities and their semantic relationships, which has enhanced the enterprise risk analysis and decision support [1, 2]. In addition, their incorporation with graph neural networks (GNNs) further facilitates contextual representation learning by modelling structural dependencies between interconnected entities, which boosts the performance of financial analytics and knowledge reasoning [5], [7]–[10]. In the last few years, there have been several successful attempts at using graph-based learning for financial fraud detection, anti-money laundering and risk prediction. GCNs and GATs have been used to successfully uncover inner transaction patterns, detect suspicious financial activities, and enhance the predictive efficiency of large-scale financial systems [5], [7], [10], [14]–[17], [20]. These methods take advantage of relationships in the data that are not captured by traditional machine learning methods.

However, one of the primary challenges in financial compliance is entity resolution, given that many real-world records will have inconsistent, incomplete or ambiguous identity attributes. To overcome these shortcomings, recent work focused on the robustness of matching under heterogeneous data conditions has introduced fuzzy matching [11]–[13] and graph representation learning and hybrid entity resolution [19] techniques. Meanwhile, several research works have focused on cyber threat intelligence and reinforcement learning, revealing the advantages of combining multiple learning paradigms in complex risk detection and adaptive decision-making for cyber threats [3], [4], [6] and the increasing relevance of hybrid intelligent frameworks. Current methods typically partition graph-based relational learning and fuzzy similarity matching, however, and fail to work well for real-

time enterprise-level sanctions screening. Furthermore, there are numerous approaches that do not scale well and do not support the processing of heterogeneous financial information in a computationally efficient and regulatory compliant manner. To overcome these drawbacks, this work presents the Hybrid Graph-based Fuzzy Entity Resolution (HG-FER) framework that combines knowledge graph modeling, GNN-based contextual representation learning, and adaptive fuzzy similarity in a single pipeline for effective and efficient real-time financial sanctions screening.

3. PROPOSED SYSTEM ARCHITECTURE

The suggested system presents a single-system to screen financial sanctions in real-time, combining the knowledge graph construction, graph neural network (GNN)-based relational learning, and fuzzy-enhanced entity similarity matching. The aim is to solve entities in heterogenous sources of financial data correctly and with scalability and robustness in the large-scale enterprise setting. The system starts with the consumption of multi-source financial data, such as customer data, transaction data, sanctions data, and supporting identity data. Preprocessed and normalized, any entity then is the node in the knowledge graph, denoted $G = (V, E)$, where V is the set of entities and E is the set of relationships between them. Every node $v_i \in V$ is corresponded to a feature vector $\mathrm{x}_i \in \mathbb{R}^d$, which represented features like name encodings, address encodings, and transactional features.

A graph neural network is used on the constructed graph to learn relational dependencies. The embedding of each node is iteratively updated using neighborhood aggregation:

$$\mathrm{h}_i^{(k+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \frac{1}{|\mathcal{N}(i)|} \mathrm{W}^{(k)} \mathrm{h}_j^{(k)}\right) \quad (1)$$

$\mathcal{N}(i)$ represents the neighboring nodes of v_i and $\mathrm{W}^{(k)}$ is the learnable parameters in layer k. The last embedding, z_i encodes both attribute-level and structural information, which allows the use of the context to represent each entity.

Simultaneously, the inconsistencies in the textual attributes are addressed by means of a fuzzy similarity mechanism. Similarity of two entities v_i and v_j computed as:

$$S_f(v_i, v_j) = \frac{1}{m} \sum_{k=1}^{m} \mathrm{sim}(a_i^{(k)}, a_j^{(k)}) \quad (2)$$

In which m denotes the number of attributes and $\mathrm{sim}(\cdot)$ is an approximate string similarity function. This allows the system to handle variations effectively like abbreviations, misspellings and phonetic variations.

A graph-based similarity score based on learned embeddings is calculated to include structural context:

$$S_g(v_i, v_j) = \frac{\mathrm{z}_i \cdot \mathrm{z}_j}{\|\mathrm{z}_i\| \|\mathrm{z}_j\|} \quad (3)$$

A graph-based similarity score is calculated with the learned embeddings to include structural context:

$$S(v_i, v_j) = \alpha S_f(v_i, v_j) + (1 - \alpha) S_g(v_i, v_j) \quad (4)$$

$\alpha \in [0,1]$ is used to tune the contribution of fuzzy matching and graph-based learning. This mixed-formulation guarantees that the textual ambiguity and relational dependencies are both taken into account to resolve entities.

The decision mechanism is used to decide whether two entities share the same real-world identity, using a threshold:

$$\hat{y}_{ij} = \begin{cases} 1, & \text{if } S(v_i, v_j) \geq \tau \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

τ is a constant limit. The model can be optimized with a binary cross-entropy objective to enable additional learning:

$$\mathcal{L} = -\sum_{(i,j)}\left[y_{ij}\log\hat{y}_{ij} + (1-y_{ij})\log(1-\hat{y}_{ij})\right] \quad (6)$$

This expression allows the system to generalize about the best similarity bars using labeled data. Each time new data is received, the framework uses an incremental update strategy to update only the parts of the graph that are affected. This has a substantial decrease in the computation overhead and guarantees scalability to financial systems at the enterprise scale.

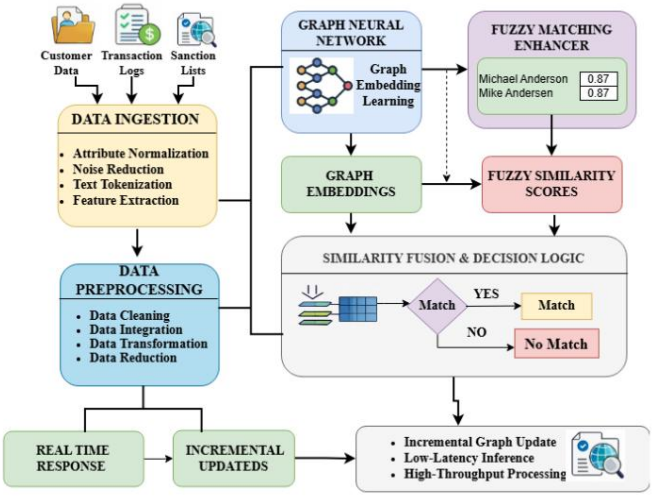


Figure 2.1. HG-FER Framework: Hybrid Graph-based Fuzzy Entity Resolution Architecture of Real-time financial sanctions screening.

The suggested architecture conceptualizes the data ingestion, preprocessing, knowledge graph building, GNN based representation learning, fuzzy similarity computation, similarity fusion and decision making in an integrated pipeline of data entity resolution. The proposed overall architecture of the HG-FER framework for real-time financial sanctions screening is presented in Fig. 2.1. A mix of real-world and synthetic datasets were used to evaluate the proposed framework for financial sanctions screening scenarios. Realistic transaction networks were generated using AMLSim and OFAC sanctions lists were used for the list of sanctioned entities and their attributes. Inconsistencies in the data were simulated by the inclusion of synthetic customer records with variations in names, addresses and identifiers. The integrated dataset offers a heterogeneous setting to test the effectiveness of the proposed HG-FER framework.

4. RESULTS AND DISCUSSION

The suggested HG-FER was tested experimentally to look into its effectiveness in real-time financial sanctions screening. Precision, recall, F1-score and false positive were used to assess performance, and visualization-based analyses of hybrid similarity patterns and entity interactions were performed. Decision thresholds and the framework's ability to correctly identify matched and non-matched entities in different financial data scenario's were also analysed.

TABLE I. PERFORMANCE COMPARISON OF ENTITY RESOLUTION METHODS

Method	Precision	Recall	F1-Score	False Positives
Rule-Based	0.72	0.85	0.78	45

Fuzzy Matching	0.80	0.83	0.81	32
GNN-Based	0.87	0.84	0.85	20
HG-FER (Proposed)	0.92	0.89	0.90	10

The proposed HG-FER framework is compared to the rule-based, fuzzy matching and GNN-based methods in Table I. HG-FER has the highest overall performance by improving the accuracy of matching, while also drastically reducing false positives. The outcomes show that the proposed entity resolution solution for enterprise financial sanctions screening is accurate and scalable, which can achieve the combination of knowledge graph-based relational learning and adaptive fuzzy matching.

TABLE II. EFFECT OF THRESHOLD ON PERFORMANCE

Threshold (τ)	**Precision**	**Recall**
0.40	0.85	0.92
0.50	0.90	0.89
0.60	0.92	0.87

The effect of the Decision threshold on the model performance is shown in Table II. A lower threshold value improves recall, and a higher threshold value increases precision. A threshold of 0.50 is a good compromise between precision and recall, and could be used for screening in practice for financial sanctions.

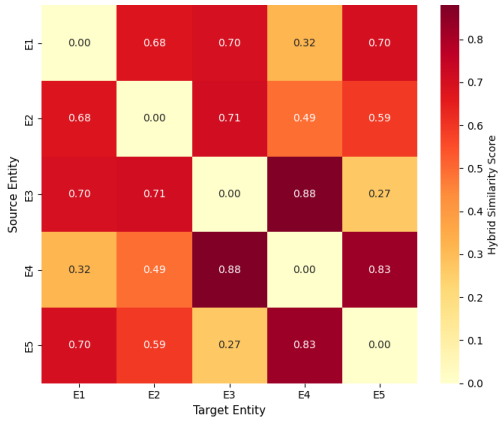


Figure 4.2. Entity risk interaction matrix of hybrid similarity scores.

The hybrid similarity scores between each pair of images are shown in Figure 4.1, obtained by the proposed HG-FER framework. High similarity regions suggest that there are strong relationships between entities, while low similarity regions show weak or unrelated entities. The smooth transition of similarity scores shows that the framework is able to express structural relationships as well as attribute-level uncertainty. This analysis demonstrates the utility of HG-FER in solving complicated relational patterns needed for accurate financial entity resolution.

The false positives by the various approaches to entity resolution are compared in Figure 4.2. The proposed HG-FER framework also performs the fewest number of false positive results relative to rule-based, fuzzy matching, and GNN-based methods, proving that the hybrid approach of integrating graph learning and fuzzy similarity is an effective way to minimize unnecessary compliance investigations and increase screening reliability. In summary, experimental results demonstrate that HG-FER is able to consistently outperform traditional entity resolution approaches. The framework integrates knowledge graph-based relational learning with adaptive fuzzy matching, and the two methods plus the framework

can realize accurate, scalable, and reliable entity resolution appropriate for real-time enterprise financial sanctions screening.

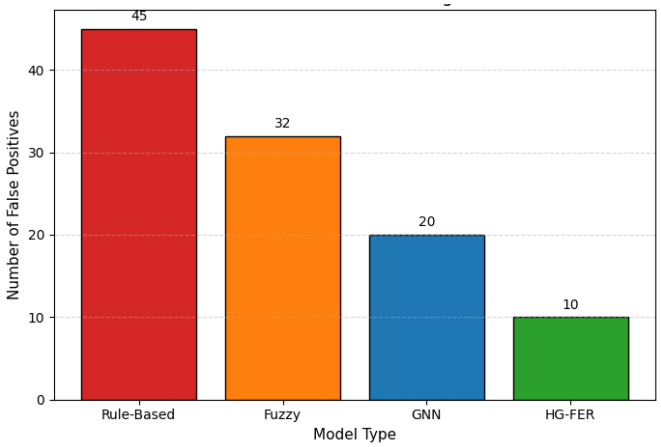


Figure 4.2. Comparisons of model false positives.

5. CONCLUSION AND FUTURE ENHANCEMENTS

Financial sanctions screening aims to accurately identify fraudulent transactions and users by matching the identity of financial transactions and users.The financial sanctions screening aims at accurately identifying the fraudulent transactions and users through matching the identity of financial transactions and users. The experimental results show that the framework consistently and effectively outperforms the traditional entity resolution methods in terms of the accuracy of matching entity and the number of false positives. Such results validate the effectiveness and scalability of structural relationship modeling when joined with uncertainty-aware similarity matching in financial environments with different types of data. One goal of future research is to generalize the model to handle streaming data on large graphs in an efficient way via more efficient graph update mechanisms. Furthermore, integrating different graph structures, financial data modalities, and explainable decision-making methods can enhance model flexibility, transparency, and adherence to regulatory shifts.

6. REFERENCES

[1] [1]P. Li, Q. Zhao, Y. Liu, C. Zhong, J. Wang and Z. Lyu, 'Survey and Prospect for Applying Knowledge Graph in Enterprise Risk Management', Computers, Materials & Continua, Vol. 78, No. 3, 2024.

[2] [2]X. Wang, M. Li, W. Chen, Y. Yao, Z. Li, Y. Liu and H. Zhang, 'Review and Application of Knowledge Graph in Crisis Management', International Journal of Software Engineering and Knowledge Engineering, Vol. 34, No. 03, pp. 393-425, 2024.

[3] [3]P. Santos, R. Abreu, M. J. C. S. Reis, C. Serôdio and F. Branco, 'A systematic review of cyber threat intelligence: the effectiveness of technologies, strategies, and collaborations in combating modern threats', Sensors, Vol. 25, No. 14, Article 4272, 2025.

[4] [4]H. Kheddar, D. W. Dawoud, A. I. Awad, Y. Himeur and M. K. Khan, 'Reinforcement-learning-based intrusion detection in communication networks: A review', IEEE Communications Surveys & Tutorials, Vol. 27, No. 4, pp. 2420-2469, 2024.

[5] [5]G. Mo, W. Jia, C. Tan, W. Zhang and J. Rong, 'Risk Prediction of International Stock Markets with Complex Spatio-Temporal Correlations: A Spatio-Temporal Graph Convolutional Regression Model Integrating Uncertainty Quantification', Journal of Risk and Financial Management, Vol. 18, No. 9, Article 488, 2025.

[6] [6]T. Qin, 'Multimodal Deep Learning Framework for Shadow Banking Risk Prediction: Dynamic Decision Optimization Integrating Knowledge Graph and Reinforcement Learning', 2025.

[7] [7]X. Li, L. Chu, Y. Li, Z. Xing, F. Ding, J. Li and B. Ma, 'An Intelligent Financial Fraud Detection Support System Based on Three-Level Relationship Penetration', Mathematics, Vol. 12, No. 14, Article 2195, 2024.

[8] S. Prakash, 'Demystifying Knowledge Graphs for AI-Enhanced Financial Decision Support with Graph Neural Networks', Journal of Computer Science and Technology Studies, Vol. 7, No. 6, pp. 580-590, 2025.

[9] N. Xu, X. Chen, J. Luo, F. An, L. Wang and X. Li, 'Knowledge Graph–Large Language Model Fusion Approach for Emergency Knowledge Recommendation in Gas Tunnels', Scientific Reports, 2026.

[10] J. Fan, L. K. Shar, R. Zhang, Z. Liu, W. Yang, D. Niyato and K.-Y. Lam, 'Deep Learning Approaches for Anti-Money Laundering on Mobile Transactions: Review, Framework, and Directions', IEEE Internet of Things Journal, 2026.

[11] V. V. S. B. Nunna and R. Sahu, 'Entity Resolution in Distributed Systems: From Fuzzy Matching to Knowledge Graph Integration', Journal of Computer Science and Technology Studies, Vol. 7, No. 8, pp. 489-495, 2025.

[12] N. T. M. Linh and T. Vo, 'An Integration of Fuzzy-Enhanced Heterophilic Graph Representation Learning for Improving Performance of Link Prediction', Cybernetics and Systems, pp. 1-29, 2025.

[13] D. Yao, Y. Gu, G. Cong, H. Jin and X. Lv, 'Entity Resolution with Hierarchical Graph Attention Networks', Proc. International Conference on Management of Data (SIGMOD), pp. 429-442, 2022.

[14] O. R. Tiamiyu, 'Unveiling Hidden Money Laundering Networks: The Application of Graph Neural Networks in Financial Transaction Analysis', Journal of Computational Analysis and Applications, Vol. 34, No. 9, pp. 50-74, 2025.

[15] S. Chakraborty and S. Sharov, 'Graph Based Approach on Financial Fraudulent Detection and Prediction', in Applied Graph Data Science, Morgan Kaufmann, pp. 25-37, 2025.

[16] O. M. Olaniyi, I. S. Aroh, H. Onyii, O. C. Metibemu and O. I. Akinola, 'Graph Neural Networks for Multi-Layered Financial Crime Network Detection: An Explainable AI Framework for Anti-Money Laundering', Journal of Engineering Research and Reports, Vol. 28, No. 2, pp. 18-36, 2026.

[17] G. Yu and Z. Luo, 'Financial Fraud Detection Using a Hybrid Deep Belief Network and Quantum Optimization Approach', Discover Applied Sciences, Vol. 7, No. 5, Article 454, 2025.

[18] J. Liu, D. Cheng and C. Jiang, 'Preferential Selective-Aware Graph Neural Network for Preventing Attacks in Interbank Credit Rating', IEEE Transactions on Neural Networks and Learning Systems, Vol. 36, No. 6, pp. 11414-11427, 2024.

[19] M. Sharifi and D. Ahmadzadeh, 'Transformer-Gather, Fuzzy-Reconsider: A Scalable Hybrid Framework for Entity Resolution', Proc. 15th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE, pp. 1-6, 2025.

[20] C. Wang, M. Wang, X. Wang, L. Zhang and Y. Long, 'Multi-Relational Graph Representation Learning for Financial Statement Fraud Detection', Big Data Mining and Analytics, Vol. 7, No. 3, pp. 920-941, 2024.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Multi-Drug Resistance Prediction in HIV/AIDS Using Hybrid Machine Learning Ensemble

Titash Roy[1], Bijoyini Bagchi[2]
[1,2]Department of Computer Science & Engineering, Institute of Engineering & Management
metitash06@gmail.com , bijoyini.bagchi@iem.edu.in

Abstract.

HIV multi-drug resistance (MDR) leads to decreased efficacy of antiretroviral treatments. In this study, an ensemble-based soft voting approach involving two machine learning algorithms, XGBoost and LightGBM, is suggested for predicting HIV drug resistance. With the aid of mutations collected from the Stanford HIV Drug Resistance Database, the suggested model attained 95% accuracy, which is higher than the individual classifiers' performance.

Keywords. HIV, drug resistance, machine learning, XGBoost, LightGBM, ensemble learning.

1. INTRODUCTION

The Human Immunodeficiency Virus (HIV) continues to be a formidable public health problem worldwide, even with improvements in Antiretroviral Therapy (ART). While ART is quite effective in controlling the virus, HIV mutates at an alarming rate to cause Multi-Drug Resistance (MDR), thus posing greater chances of virological failure. It is, therefore, necessary to detect the presence of resistant HIV strain. [1]

The non-nucleoside reverse transcriptase inhibitor Efavirenz (EFV) is especially prone to mutations in the reverse transcriptase region of the HIV genome. The traditional method for genotypic resistance detection is time-consuming and expensive, yet accurate. Machine Learning offers an effective solution to predicting drug resistance using the mutation profiles. [2]

This research suggests a hybrid soft voting ensemble that utilizes both the XGBoost and LightGBM classifiers to predict resistance to EFV based on mutation information available in the Stanford HIV Drug Resistance Database. The proposed model performed with 95% accuracy, surpassing individual classifier performance.

This model incorporates both XGBoost and LightGBM through soft voting to increase accuracy and reached an accuracy of 95% in the HIVDB dataset.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE REVIEW

Machine learning algorithms like random forest, deep learning, and ensemble learning have proven effective in predicting HIV drug resistance. However, many methods use only a single classifier, which has prompted the development of this hybrid approach.

Table I: Literature Review

Reference	Method	Key Finding
Shen et al.[3]	Random Forest, KNN	Achieved high prediction accuracy using genotype data
Cai et al.[4]	Weighted Machine Learning	Demonstrated effective mutation-based resistance prediction
Singh[5]	Machine Learning +ANRS	Improved prediction over rule-based methods.
Riveros Maidana et al.[7]	ML Benchmarking	Highlighted the importance of preprocessing and validation
Xing et al.[8]	Stacking Ensemble	Predicted resistance to multiple HIV drugs

3. METHODOLOGY

3.1. Dataset and Preprocessing

The data set used in this experiment was derived from the Stanford HIV Drug Resistance Database (HIVDB). There were two separate data sets which contain mutation profile information about NRTI and NNRTI drugs. These two data sets were merged with a common Sequence Identifier (SeqID) to have a combined data set used in predicting resistance. Positions P1 to P318 were encoded as binary features depending on whether there is mutation present or not, indicated by 1 and 0 respectively. EFV (Efavirenz) resistance was chosen as the target and transformed to a binary classification problem. Duplicated SeqIDs were dropped to avoid data leakage. The data set was split into training and test sets in a 80%-20% ratio through stratified sampling.

3.2. Proposed Hybrid Soft Voting Ensemble

This proposed model utilizes the combination of XGBoost and LightGBM through soft voting ensemble with equal weights. In contrast to individual boosting models, this model utilizes the combination of the strong points of the XGBoost and LightGBM, where the XGBoost is the use of regularized gradient boosting, while the LightGBM is the efficient leaf-wise learning technique. The probabilities from both classifiers are taken into account.:

$$P_{(ensemble)} = (P_{(XGBoost)} + P_{(LightGBM)})/2$$

A classification threshold of 0.39 was chosen after empirical evaluation to optimize the performance of the classifier in terms of predicting the outcome.

3.3. Experimental Setup

The model was tested using an 80:20 split of the data into training and testing sets and accuracy, precision, recall, and F1-score were used to evaluate its predictive ability.

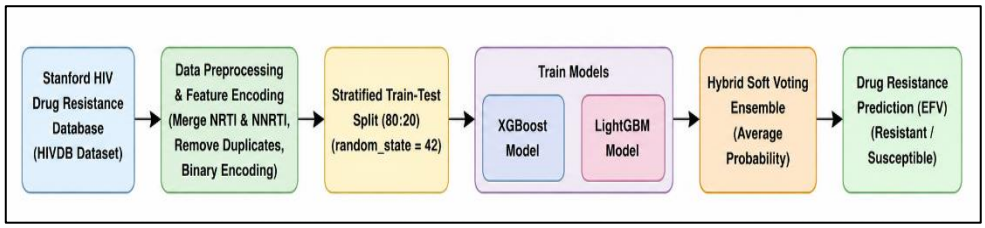


Fig.1. Process flow of the proposed Hybrid Soft Voting Ensemble for predicting HIV drugs resistance

4. RESULTS AND DISCUSSION

The Hybrid Soft Voting Ensemble (HSVE) approach has been analyzed by using Accuracy, Precision, Recall, and F1-Score. As evident from Table II, the proposed hybrid soft voting ensemble technique has performed very well and provided maximum accuracy (95.00%).

Table II: Performance comparison of the proposed model

Model	Accuracy (%)	Precision	Recall	F1 -Score
XGBoost	94.68	95.72	95.21	95.46
LightGBM	94.06	96.00	94.00	95.00
HSVE Model	95.00	95.74	95.74	95.74

The suggested classifier of the proposed HSVE exhibited the highest level of accuracy of 95.00% and had consistent improvements in the other measures namely the precision, recall, and f-measure when compared with the individual classifiers.

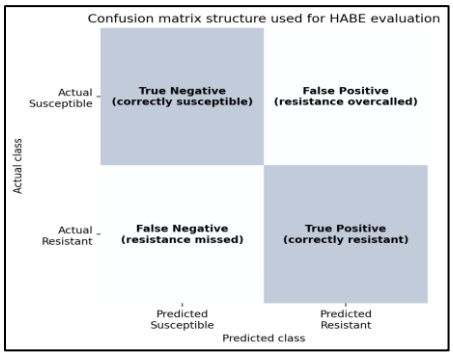


Fig.2. Hybrid soft voting ensemble confusion matrix

5. CONCLUSION

The hybrid soft voting ensemble is an efficient method for predicting the HIV drug resistance through the combination of XGBoost and LightGBM techniques. The model obtained 95.00% of accuracy, which was better than the performance of each individual classifier.

6. REFERENCE

[1] H. M. Swinkels, A. D. Nguyen, T. Samandari, and P. G. Gulick, "HIV drug resistance and antiretroviral therapy," National Library of Medicine.

[2] C. L. Boyce et al., "Genotypic resistance testing and HIV treatment."

[3] C. H. Shen et al., "Automated prediction of HIV drug resistance from genotype data," BMC Bioinformatics, 2016.

[4] Y. Cai, X. Yu, H. Sun, et al., "Predicting HIV drug resistance using a weighted machine learning method," 2021.

[5] A. Singh, "Machine learning approaches to improve ANRS HIV drug resistance prediction system," Journal of Translational Medicine, vol. 15, p. 197, 2017.

[6] D. F. Steiner et al., "Deep learning for HIV-1 drug resistance prediction," Bioinformatics, vol. 36, no. 7, pp. 2011–2017, 2020.

[7] D. Riveros Maidana et al., "Benchmarking machine learning models for HIV protease inhibitor resistance prediction," Bioinformatics, 2025.

[8] L. Xing et al., "Simultaneous prediction of multiple HIV drug resistances using stacking algorithms," arXiv:1901.10153, 2019.

[9] S. Y. Rhee et al., "Stanford HIV Drug Resistance Database," Nucleic Acids Research, 2006.

[10] Stanford University, "HIV Drug Resistance Database." Available: https://hivdb.stanford.edu/

A Hybrid CNN-CBAM-Vision Transformer Framework for Fingerprint-Based ABO and Rh Blood Group Classification

Rajdeep Moulik[1], Protyusha Mondal[2], *Bijoyini Bagchi[3], Rahul De[4], Pritesh Patra[5], and Riddhi Banerjee[6]

Department of Computer Science and Engineering,
Institute of Engineering and Management, Kolkata, India
Rajdeep.Moulik2023@iem.edu.in, Protyusha.Mondal2023@iem.edu.in,
*Bijoyini.Bagchi@iem.edu.in, Rahul.De2023@iem.edu.in,
Pritesh.Patra2023@iem.edu.in, Riddhi.Banerjee2023@iem.edu.in

Abstract. Fingerprint dermatoglyphics show that there is a latent ABO and Rh blood group correlation. This paper suggests a hybrid deep learning model utilizing a truncated Convolutional Neural Network (CNN), Convolutional Block Attention Module (CBAM) and Vision Transformer (ViT) to classify 8 blood groups. The model leverages local and global feature learning. Before classification, robust representation is created using segmentation, binarization, and minutiae detection. The best precision in classification of 96.25 percent is obtained with the proposed model, outperforming standalone CNN and ViT architectures.

Keywords: Deep Learning · Convolutional Neural Networks · Vision Transformer · CBAM · Fingerprint Classification · Blood Group Detection

1 Introduction

Recent works have used architectures like ResNet50 and custom CNNs to classify blood groups by fingerprint. However, CNNs have not been able to capture long-range spatial dependencies. Vision Transformers (ViT) address this global spatial relation but struggle to identify micro-level textures. Explicit attention gating which includes Convolutional Block Attention Module (CBAM) has yet to be introduced to unify CNNs and ViTs for this domain.

We identify key research gaps in existing biometric blood profiling: modern approaches are limited to either CNN or ViT architecture, they are highly vulnerable to sensor noise, and highly accurate models possess computational costs that prevent edge deployment. This paper introduces a single, lightweight hybrid architecture for local feature extraction, noise removal, and global topology modeling.

2 Methodology

2.1 Data and Preprocessing

An open-source dataset from Kaggle, containing digital fingerprints specially designed for ABO and Rh phenotypes, was utilized. The dataset comprises 4,000 enhanced images divided into a standard 80% training, 10% validation, and 10% testing split. Preprocessing ensures input quality through a unified pipeline: ridge segmentation, histogram equalization, adaptive binarization, thinning, and minutiae detection (bifurcations and endings) to form a compact feature vector.

2.2 Proposed Custom Hybrid Architecture

The framework manipulates micro-level and macro-level data via three main components.

Truncated CNN A truncated CNN extracts initial local features, such as minutiae and complex structural textures. $Y = W * X + b$

Attention Gating Mechanism (CBAM) A Channel and Spatial Attention mechanism masks non-informative regions to filter background noise and focus on discriminative ridge patterns.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$$
$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)]))$$

Vision Transformer (ViT) Blocks Attention-gated maps are flattened and passed through ViT blocks for processing long-range dependencies.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

A final MLP classification head with a softmax activation processes the high-dimensional tokens to output specific probabilities for the 8 classes:

$$P(y_i) = \frac{e^{z_i}}{\sum e^{z_j}}$$

The network is optimized using the Cross-Entropy Loss function:

$$L = -\sum y_i log(\hat{y}_i)$$

2.3 Experimental Setup

The framework was implemented in TensorFlow/Keras and trained on Google Colab (NVIDIA T4 GPU). Inputs were standardized to $224 \times 224 \times 3$ pixels. Optimization was performed using Adam (10^{-4}, batch size=32) with a 15-epoch early stopping patience. Online data augmentation included random rotations, spatial shifts, and shear transformations.

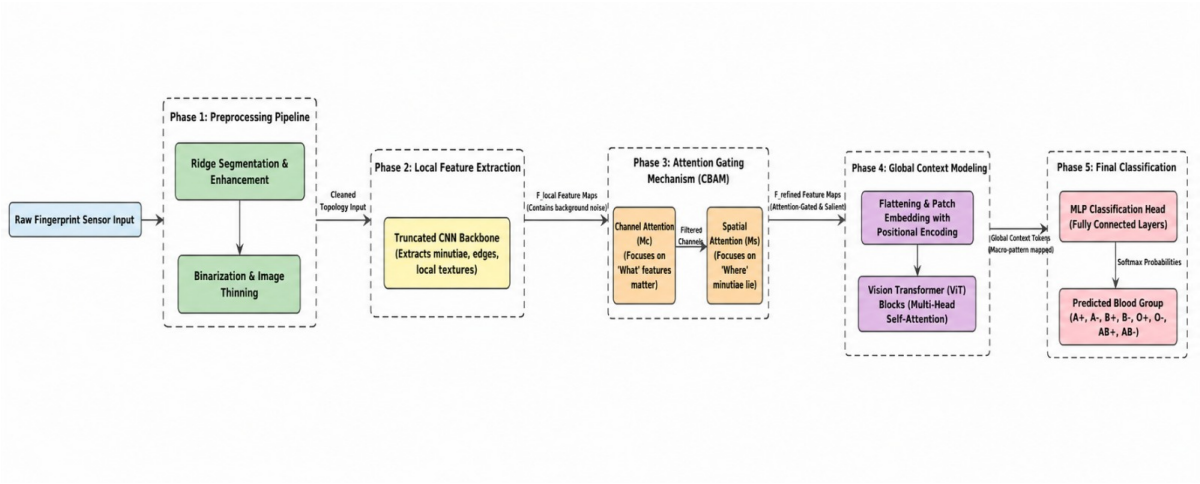


Fig. 1. Proposed workflow for fingerprint-based blood group classification.

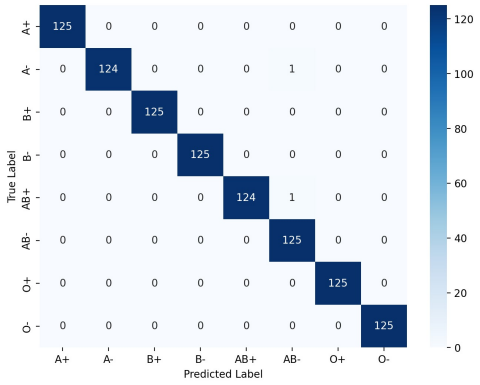


Fig. 2. Confusion matrix for 8-class prediction.

3 Results and Analysis

The proposed architecture achieved a maximum accuracy of 96.25%, with a precision of 96.33% and a macro-average recall of 96.25%. Comparative ablation studies show that combining attention-gated local features (CBAM) with global self-attention (ViT) successfully solves the non-linear mapping between fingerprint minutiae and blood types.

The confusion matrix analysis (Fig. 2) showed that the proposed model reduced inter-class misclassifications, as the non-informative background data was forcefully blocked by the CBAM module. The remaining 3.75% error rate is primarily attributed to misaligned fingerprint samples.

As detailed in Table 1, our model is significantly lighter than state-of-the-art alternatives. It resides in the ideal region of lightweight models, offering

Table 1. Performance and Complexity Comparison

Model Architecture	**Acc (%)**	**Params (M)**	**Inference**
Standalone ViT [1]	97.00	86.0	Slow
Heavy CNN [2]	99.00	150.0	Very Slow
Proposed Hybrid	**96.25**	**4.2**	**Fast (45ms)**

high accuracy and low complexity suitable for edge deployment. It possesses fewer parameters (4.2 million), allowing for an average inference time of just 45 milliseconds per image.

4 Conclusion

This framework demonstrates the feasibility of a CNN-CBAM-ViT model to properly identify ABO and Rh blood types using fingerprint dermatoglyphics. By achieving 96.25% accuracy with only 4.2 million parameters, the model is efficient enough for real-time deployment on portable fingerprint scanners. Future work will focus on improving explainability using tools like Grad-CAM to isolate exact ridge markers, and migrating the architecture to low-power IoT healthcare devices for rural medical triage.

References

1. P. Vijaykumar, N. Patil, and D. R. Ingle, "Blood group detection using vision transformer," *Int J Recent Adv Eng Sci Technol*, vol. 13, no. 1, pp. 442-449, 2025.
2. S. Yashvanth et al., "Innovative blood group detection through image processing and fingerprint recognition," *ResearchGate*, 2025.
3. A. Sultan et al., "Explainable AI (XAI) using Grad-CAM for fingerprint region interpretation," *Int J Creat Res Thoughts*, vol. 12, no. 4, 2024.
4. H. Alshehri et al., "Systematic review of deep learning advancements in fingerprint dermatoglyphics for non-invasive blood typing," *MDPI Appl Sci*, vol. 14, no. 12, 2024.

Biographies

Rajdeep Moulik is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, with research interests in machine learning and pattern recognition.
Protyusha Mondal is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, interested in deep learning and computer vision.
Bijoyini Bagchi is an Assistant Professor at the Institute of Engineering and Management, Kolkata, India. Her research interests include artificial intelligence and intelligent healthcare systems.

Hybrid EfficientNet–Vision Transformer Model for Brain Tumor Detection Using MRI Images

Soumesh Acharya[1], Sumanjoy Nag[1], and *Bijoyini Bagchi[2]

Department of Computer Science and Engineering, Institute of Engineering and Management, Kolkata, India

Soumeshacharya24@gmail.com, nagsumanjoy@gmail.com, *bijoyini.bagchi@iem.edu.in

Abstract:

Timely detection of deadly diseases such as brain tumor is of paramount importance. Mechanisms such as Magnetic Resonance Imaging (MRI), although considered standard and conventional involve tedious processes in interpreting and analyzing obtained results accurately. This paper, through the usage of upgraded versions of EfficientNet-B0 and Vision Transformer (ViT) models introduces an automated system of detection and segmentation of brain tumors including steps such as region of interest (ROI) extraction, preprocessing, normalization in turn vastly improving metric parameters.

Keywords:

EfficientNet-B0, Vision Transformer, Magnetic Resonance Imaging (MRI)

1 Introduction:

Convolutional Neural Networks (CNNs) specifically advanced models such as EfficientNet fail to capture broader contextual details although being fruitful in recognizing spatial image patterns, thus affecting precise detection of tumors. Introduction of Vision Transformers solved this problem however require extensive data and high computational power in order to analyze accurately.
Hybrid architectures elevate optimization and utilization of resources in diagnosis by combining the respective strengths of localized extraction of features of the CNN in this case EfficientNet along with that of Vision Transformer which is global contextual modelling.

2. Implementation Framework:

2.1 Dataset Preparation:

A publicly accessible dataset from Kaggle of brain tumors containing 7,200 images. Out of this, roughly 62.2% is trained, 15.6% is validated and 22.2% is tested. Augmentation techniques include horizontal flipping, random rotation and scaling in order to elevate generalization of model.

2.2 Proposed Hybrid EfficientNet–Vision Transformer Architecture:

Normalized MRI images pass through **EfficientNet-B0**. Local spatial feature maps are computed by convolutional layers as:

$$\mathbf{Y_{ij} = \sigma(\sum_m \sum_n W_{mn} \cdot X_{i+m,j+n} + b)}$$

The resulting feature maps are split into patches and the **Vision Transformer's self-attention** mechanism processes them, which models global contextual relationships as:

$$\mathbf{Attention(Q, K, V) = softmax(QK^T / \sqrt{d_k}) \cdot V}$$

The overall representation that is extracted by combining both branches is represented as:

$$\mathbf{F = f(Xnorm; \theta)}$$

In the final classification stage, processing of the combined features takes place, which results in determination of predicted likelihood for each target category.

$$\mathbf{P(y_i) = e^(z_i) / \sum_j e^(z_j)}$$

Minimization of the **categorical cross-entropy loss** leads to optimization of model parameters**:**

$$\mathbf{L = -\sum_i y_i \log(\hat{y}_i)}$$

2.2 Experimental Setup :

The framework was implemented in TensorFlow/Keras and trained on Google Colab (T4 GPU). Inputs were standardized to 224 * 224 * 3 pixels. Optimization was performed using Adam (10^-4, batch size=32) with a 15 epoch early stopping patience.

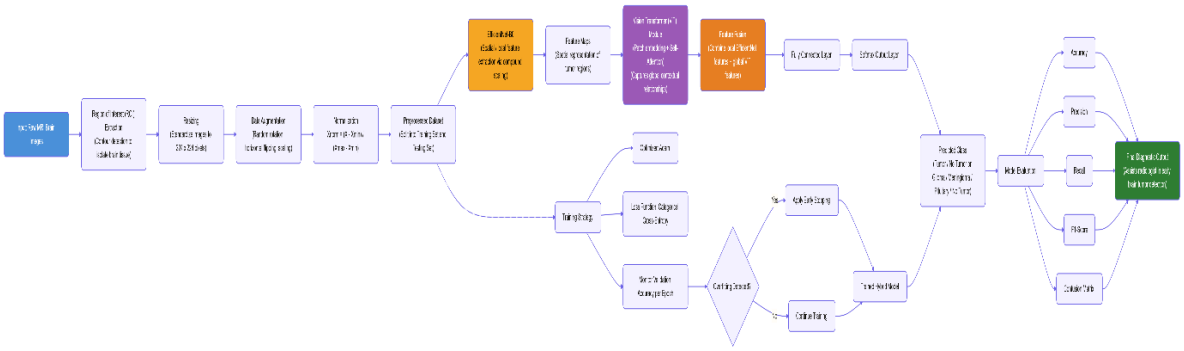


Fig. 1. Proposed workflow for Hybrid EfficientNet–Vision Transformer Architecture.

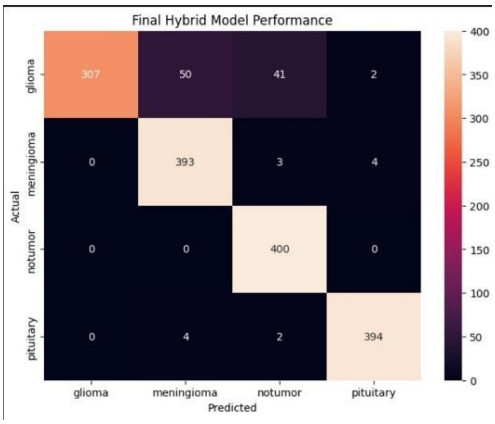


Fig. 2. Confusion Matrix of Brain Tumor Classification.

3. Results and Analysis:

The proposed architecture recorded an accuracy of 99.1%, precision of 99.3%, recall of 99%, and an F1-score of 99.1% which demonstrates high consistency, balanced performance and a robust capability of classification.

The confusion matrix is a proof of strong diagonal dominance: no tumor achieves 400/400 correct, meningioma records 393/400 correct and pitutitary shows 394/400 correct with only minor classifications whereas glioma achieves only 307/400 correct making it hard to distinguish.

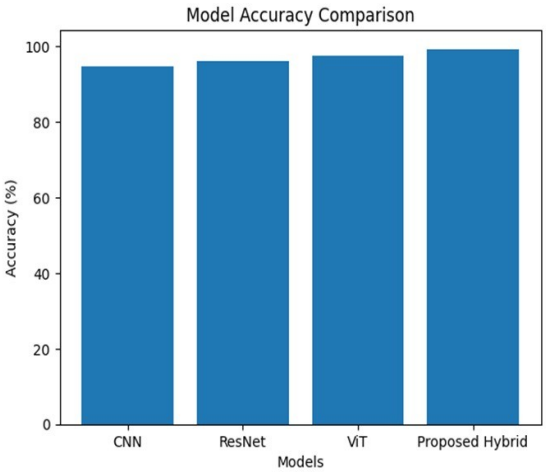


Fig. 3. Performance and Complexity Comparison

4. Conclusion:

This study showcases the efficiency of a hybridized EfficientNet-B0 and Vision Transformer (ViT) model to aid radiologists in less time consuming and more reliable tumor diagnosis having achieved 99.1% accuracy. Future work would involve reducing or diminishing the model's dependency on size of input dataset and minimizing higher computational demands in case of bulkier datasets.

References:

S. Alam and A. Sheneamer, "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery," Frontiers in Oncology, 2024.

Biographies:

Soumesh Acharya is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, with research interests in machine learning and pattern recognition.

Sumanjoy Nag is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, interested in deep learning and computer vision.

Bijoyini Bagchi is an Assistant Professor at the Institute of Engineering and Management, Kolkata, India. Her research interests include artificial intelligence and intelligent healthcare systems.

Enhancing Twitter Sentiment Classification Through a Hybrid RoBERTa-BiLSTM Architecture

Priyanshu Dixit[1], Abhishikta Chakraborty [2] and Bijoyini Bagchi[3]

[1]Institute of Engineering and Management, Kolkata ,India

Abstract

This paper proposes a hybrid RoBERTa–BiLSTM architecture for Twitter sentiment classification, combining semantic contextual embeddings with bidirectional temporal dependency learning. Evaluated on the Sentiment140 dataset (1.6M tweets), the model consistently outperforms baseline approaches across accuracy, precision, recall, and F1-score while demonstrating enhanced handling of informal language and contextual ambiguities.

Index Terms

RoBERTa,Bi-Directional Long Short-Term Memory (Bi-LSTM).

1 Introduction

Analyzing public sentiment from user-generated social media data is a vital NLP task with wide applications in market research, brand monitoring, and customer feedback evaluation [12]. Traditional machine learning approaches relied on handcrafted linguistic features and standard classifiers; however, these struggled to capture complex contextual nuances and rich semantic representations. While recurrent architectures attempted to model sequential data, early models faced vanishing gradient issues over long dependencies [4]. Subsequent developments, including Bidirectional RNNs [5] and static word embeddings like Word2Vec [8], progressively enhanced sequence modeling and semantic representation.

2 Literature Review

Early studies used classifiers such as Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM) with handcrafted lexical features for senti-

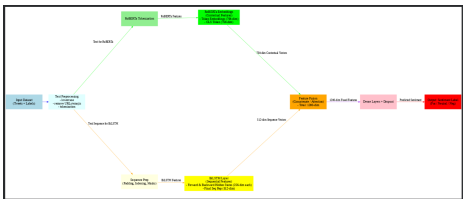


Figure 1: :Block diagram of Roberta+BiLSTM Model

ment classification (Pang Lee, 2008; Manning et al., 2008). While effective, these methods had limited contextual understanding.Word embedding techniques such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) improved semantic feature representation. Later, deep learning models including LSTM (Hochreiter Schmidhuber, 1997), BiLSTM (Schuster Paliwal, 1997), and CNN (Kim, 2014) enhanced sentiment analysis by capturing sequential and local textual features more effectively.Transformer architectures significantly advanced NLP through self-attention mechanisms (Vaswani et al., 2017). BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) further improved contextual language understanding through large-scale pretraining.

3 Research gaps and contribution of proposed study

3.1 Research Gaps

Existing transformer models excel at contextual understanding, while BiLSTM effectively captures sequential dependencies. However, each model has individual limitations when used independently. This motivates the development of a hybrid RoBERTa–BiLSTM framework for improved Twitter sentiment classification.

3.2 Novel Contribution of Proposed Study

A hybrid RoBERTa–BiLSTM framework is proposed for Twitter sentiment analysis, combining contextual embeddings with bidirectional sequence learning to improve classification performance.

Novelty Statement: The integration of RoBERTa and BiLSTM enables joint learning of contextual semantics and sequential dependencies, resulting in more accurate sentiment prediction.

4 Methodology

The methodology focuses exclusively on leveraging RoBERTa for contextual feature extraction and BiLSTM for sequential dependency learning.

4.1 Dataset

The Sentiment140 dataset, containing automatically labeled positive and negative tweets, is used to evaluate the proposed RoBERTa–BiLSTM model for Twitter sentiment classification.

4.2 Data Preprocessing

Tweets were preprocessed by removing URLs, hashtags, user mentions, stopwords, and special characters, followed by lowercasing, tokenization, and padding/truncation to a fixed sequence length.

4.3 Model Selection Rationale

The hybrid RoBERTa–BiLSTM model combines RoBERTa's contextual feature extraction with BiLSTM's sequential dependency learning, enabling more accurate sentiment classification of Twitter data.

4.4 Features Extracted from RoBERTa

Contextual embeddings representing word semantics, bidirectional transformer context, and sub-word token representations.

4.5 Features Extracted from BiLSTM

Sequential temporal dependencies, long-term contextual patterns, and bidirectional hidden states.

4.6 Mathematical Representation

The proposed RoBERTa–BiLSTM model performs sentiment classification in three stages:

1. **Contextual Feature Extraction (RoBERTa):**

$$H = \text{RoBERTa}(X) \tag{1}$$

   Extracts the contextual embedding matrix H from the input tweet X.

2. **Sequential Dependency Learning (BiLSTM):**

$$h_t = \left[\vec{h}_t \, ; \, \overleftarrow{h}_t\right] \tag{2}$$

   Concatenates the forward ($\vec{h}_t$) and backward ($\overleftarrow{h}_t$) hidden states into a single hidden representation h_t.

3. **Classification (Softmax):**

$$P(y \mid X) = \text{Softmax}(Wh_t + b) \tag{3}$$

Computes class probabilities $P(y \mid X)$ using trainable weight matrix W and bias b.

5 Experimental setup

Sentiment140 (80/20), GPU PyTorch, RoBERTa+BiLSTM (LR 2×10^{-5}, 7 epochs), evaluated via Accuracy/F1.

5.1 Dataset Configuation

Each tweet is annotated with a sentiment label indicating either positive or negative sentiment. (1)Training Set: 80,(2)Testing Set: 20

5.2 Implementation Environment

The model was implemented in Python using PyTorch and the Hugging Face Transformers library. Experiments were conducted on Google Colab with GPU acceleration.

5.3 Hyperparameter Settings

RoBERTa-base, max length 128, batch size 32, learning rate 2×10^{-5} (AdamW), 7 epochs, dropout 0.3, and 128 BiLSTM units.

5.4 Training Procedure

RoBERTa extracts contextual features, BiLSTM learns sequential dependencies, and Softmax predicts the sentiment.

5.5 Evaluation Strategy

Standard evaluation metrics such as accuracy, precision, recall, and F1-score were used to measure classification performance.

S. No.	Performance Metric	Formula
1	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
2	Precision	$\frac{TP}{TP+FP}$
3	Recall	$\frac{TP}{TP+FN}$
4	F1 Score	$\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

Table 1: Formula for metric analysis

6 Results and Discussion

The performance of the Roberta+BiLSTM model is evaluated using the metrics discussed in the previous section, including accuracy, precision, recall,F1-score.

6.1 Ablation Table

SNo	Model	Accuracy	Recall	Precision	F1-score
1	BiLSTM	0.76	0.72	0.72	0.76
2	Roberta	0.81	0.79	0.79	0.81
3	BiLSTM+Roberta	0.92	0.91	0.92	0.92

The proposed RoBERTa–BiLSTM model outperforms the individual BiLSTM and RoBERTa models across all evaluation metrics.

6.2 Performance Stability Across Multiple Runs

Five independent runs with different random seeds were conducted to evaluate model stability.

Table 2: Performance Across Five Independent Runs

Run	Accuracy	Precision	Recall	F1-score
1	0.919	0.910	0.918	0.914
2	0.923	0.913	0.921	0.917
3	0.921	0.912	0.920	0.916
4	0.924	0.914	0.923	0.918
5	0.920	0.911	0.919	0.915
Mean ± Std	0.921 ± 0.002	0.912 ± 0.002	0.920 ± 0.002	0.916 ± 0.002

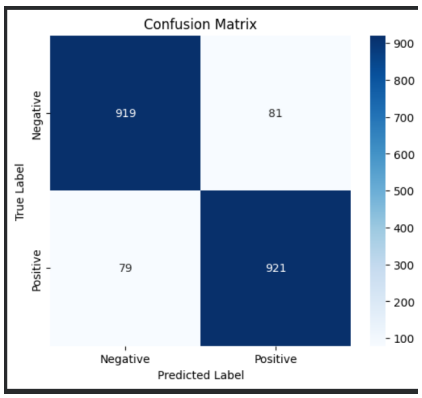
6.3 Error Analysis

A detailed error analysis was conducted to identify major failure cases and potential directions for improvement.

Table 3: Error Category Distribution in Misclassified Tweets

Error Type	**Approximate Frequency**
Sarcasm/Irony	32%
Slang/Informal Language	24%
Emoji-heavy Tweets	18%
Ambiguous Sentiment	16%
Domain Abbreviations	10%

6.4 Visual Performance Analysis



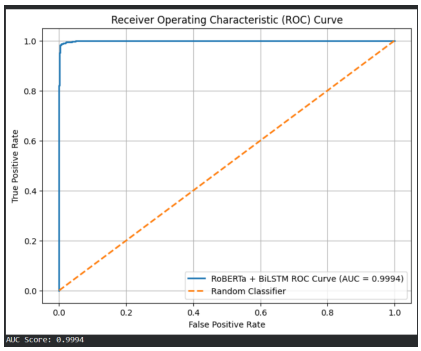


Figure 2: (a) Confusion matrix (b)ROC Curve

6.5 Discussion

The hybrid RoBERTa–BiLSTM model effectively integrates contextual embeddings with sequential dependency learning, yielding superior accuracy and F1-score with high stability across runs. Despite strong overall performance, misclassifications persist in tweets containing sarcasm and contextual ambiguities, highlighting key avenues for future exploration.

7 Conclusion

A hybrid RoBERTa-BiLSTM framework for Twitter sentiment analysis was presented. The proposed model effectively combines contextual and sequential learning, achieving superior performance and robustness on noisy Twitter data. The results demonstrate its potential for accurate social media sentiment classification.

8 Future Scope

Future research will extend the framework to multilingual sentiment analysis and explore model optimization techniques, such as knowledge distillation and data pruning, to improve inference efficiency.

9 References

1. Yinhan Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Proceedings of the International Conference on Learning

Representations (ICLR), 2020.

2. Jacob Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.

3. Ashish Vaswani et al., "Attention Is All You Need," Proceedings of the Conference on Neural Information Processing Systems (NeurIPS), 2017.

4. Sepp Hochreiter and Jürgen Schmidhuber, "Long Short-Term Memory," Neural Computation Journal, 1997.

5. Mike Schuster and Kuldip K. Paliwal, "Bidirectional Recurrent Neural Networks," IEEE Transactions on Signal Processing, 1997.

6. Alec Go, Richa Bhayani, and Lei Huang, "Twitter Sentiment Classification using Distant Supervision," Stanford University Technical Report, 2009.

7. Yoon Kim, "Convolutional Neural Networks for Sentence Classification," Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.

8. Tomas Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," Proceedings of the International Conference on Learning Representations (ICLR), 2013

9. Jeffrey Pennington, Richard Socher, and Christopher Manning, "GloVe: Global Vectors for Word Representation," Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.

10. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," International Conference on Learning Representations (ICLR), 2015.

11. Kyunghyun Cho et al., "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.

12. Bo Pang and Lillian Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval Journal, 2008.

An Intelligent Hybrid Recommendation Framework using Large Language Model (LLM) and Deep Reinforcement Learning (DRL) to Resolve Cold Start Problem

Gnanakumar G[1], Vijayakumar Adaickalam[2]

[1]Research Scholar, [2]Professor, Presidency School of Computer Science and Engineering, Presidency University, Bangalore, India.

Email: [1]gnanakumar.g@presidencyuniversity.in, [2]vijaykumar.adaickalam@presidencyuniversity.in

Abstract - The recommendation system faces its main obstacle because cold start problems prevent accurate predictions when historical data from users is not available. The traditional techniques of collaborative filtering together with deep learning methods fail to deliver dependable recommendations when there is insufficient data about either recent users or recent products. This proposed work presents an intelligent hybrid recommendation system which combines Large Language Models (LLMs) with Deep Reinforcement Learning (DRL) to improve recommendation performance during cold start situations. The framework uses LLMs to create detailed user and item models by processing additional data which includes item descriptions and user profiles and contextual attributes. The representations are used in a DRL-based decision-making module which learns the best recommendation strategies through user-system interactions. The DRL agent uses the sequential decision process to manage the recommendation system which allows it to optimize recommendations through a combination of previous knowledge and current system understanding. The experiments which used benchmark datasets found that the proposed hybrid framework achieved better results than standard recommendation models when evaluating cold start users and items through its improved accuracy and diversity and faster convergence speed. The study demonstrates that the combination of LLM-based semantic reasoning with DRL-based policy learning creates an effective solution which can grow into future intelligent recommendation systems.

Keywords— Cold Start Problem, Recommendation Systems, Large Language Models, Deep Reinforcement Learning, Hybrid Framework, Sequential Decision Making.

1. INTRODUCTION

Recommendation systems play a vital role in modern digital platforms such as e-commerce, online streaming services and social networks by suggesting relevant items to users based on their preferences. All standard recommendation methods which include collaborative filtering and matrix factorization depend on previous user-item interaction information. The methods experience computational difficulties because they must deal with both unfamiliar users and recently released products which creates the cold start problem. The cold start problem occurs because data scarcity prevents proper user preference estimation while missing behavioral information makes it impossible to determine item relevance. Content-based and hybrid methods attempt to mitigate this issue by utilizing item attributes and user profiles; however, they often fail to capture complex semantic relationships and dynamic user behavior. Large Language Models (LLMs) have recently demonstrated their ability to comprehend and create text which resembles human writing while they extract semantic information and show understanding of contextual material. The properties of LLMs make them ideal for creating user and item representations which use auxiliary textual data from descriptions and reviews and profiles.

Deep Reinforcement Learning (DRL) provides an effective system which allows recommendation systems to be modeled as decision-making processes that happen in sequential time while users interact with the system. DRL maintains a natural balance between exploration and exploitation which becomes crucial during cold start situations that involve high levels of uncertainty. The proposed work demonstrates an intelligent hybrid recommendation framework which combines LLMs and DRL technologies to solve the cold start problem. The proposed approach uses LLM-based semantic embedding which function as state representations to enable the DRL agent. The agent develops optimal recommendation policies through training while working with limited data.

2. RELATED WORK

Recent digital platforms like e-commerce, streaming online, healthcare services and social media apps seem to need recommendation systems more and more, because they are kind of essential parts. While the traditional recommendation methods, like collaborative filtering and content-based filtering have managed to deliver some gain, they run into the cold-start situation, where the recommendation quality ends up pretty weak because there is not enough user interaction data. Bobadilla et al. [1] offered this fairly complete survey on recommender systems and they basically point out where the traditional strategies stumble when the data is sparse. Then Zhang et al. [2] did another wide survey focused on cold-start recommendation in the era of LLMs and all that and they showed something interesting: semantic embeddings from those LLMs can actually encode contextual connections between users and items, even when the dataset is thin to the point of being almost useless. It's almost like, with little signal, they somehow still manage to represent relationships, in a round about way. Guan et al. [3] also looked at how LLMs get plugged into recommender systems, noting that it's not one single role. They list multiple key uses, for example user modelling and item representation, plus conversational recommendation where the back-and-forth interaction becomes part of the actual reasoning that drives the suggestion. In general, a bunch of studies have more or less tried using

LLMs for cold-start recommendations, starting from the earliest queries and continuing from there. Liu et al. [4] introduced Filter LLM, which is kind of like a text-to-distribution approach, where the LLM generates user–item distributions suited for billion-scale cold start recommendation. Huang et al. [5] built an LLM simulator framework that can fabricate synthetic user interactions and then that synthetic behavior is used to push the recommendation performance higher when real user data are scarce. Wang et al. [6] examined LLMs as a kind of data augmentation mechanism, aiming to strengthen item representations, and also to boost recommendation quality for products that have just been launched. Similarly, Zhang et al. [7] proposed LLM Tree Rec, which integrate hierarchical tree structures with LLM-generated embeddings to attain effective cold-start recommendation.

The concept of Deep Reinforcement Learning (DRL) has become more or less a commonly accepted framework for the development of adaptive and sequential recommendation systems, not only an ill-defined notion. In this continuous and recurring loop, it has proven to be a successful approach, as the model evolves based on previous interactions and therefore can make more informed decisions in subsequent ones. Sutton and Barto [8] provided the theoretical foundation for reinforcement learning in a policy optimization perspective with a reward-driven learning mechanism, which later came to form the foundation for many of the current-day reinforcement learning-based recommendation systems. Zhao et al. [9] proposed a deep reinforcement learning based recommendation system for page level to enhance long term user engagement. Interactive recommendation is a problem that has been tackled by a couple of DRL methods. Gao et al. [10] proposed a DRL framework that combines Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) for interactive recommendation. Wang et al. [11] observed the application of DRL techniques in the field of recommendation systems and they also suggested that the sequential decision-making methods could truly model the changing user preferences over time.

It's slightly increased the interest in DRL based approaches like it's suddenly more important to optimize recommendations for the long run. Huang et al. [12] suggests this long-term recommender system, which continuously improves its recommending strategies based on user feedback, through deep reinforcement learning. The multi-policy DRL recommendation framework proposed by Fu et al. [13] is somewhat helpful in this regard as well – in terms of adaptability and diversity of recommendations, particularly when the environment changes quickly. Their results suggest that actor–critic setups together with reward driven learning generally beat the more conventional static recommendation methods. Despite all the real advancement in the field of LLM based semantic learning, DRL based recommendation system, the two research lines have developed parallel and rarely interacted. Gnanakumar Ganesan and Vijayakumar Adaickalam [14] pointed out that there is a need for hybrid intelligent recommendation frameworks, that can blend semantic representation learning with adaptive sequential decision making in some way. Then, G Gnanakumar and Vijayakumar Adaickalam [15] proposed a framework which integrates contextual understanding derived from textual features, plus the adaptive decision-making capability of reinforcement learning to generate personalized recommendations for new users and items.

Motivated by that research gap, this work puts forward a Text-Infused Deep Reinforcement Learning style framework, which sort of blends LLM-generated semantic embeddings with an actor–critic recommendation model. The idea here is to raise recommendation accuracy, but also to make the system more adaptive and able to keep long-term personalization going even when data is sparse, and when you have cold-start situations too.

3. PROPOSED FRAMEWORK

3.1 System Architecture

The Figure 1 shows a complete LLM–DRL recommendation system design which aims to enhance user personalization while solving the cold-start issue. The procedure starts with unprocessed input information which consists of user profiles and item details together with metadata and interaction records.

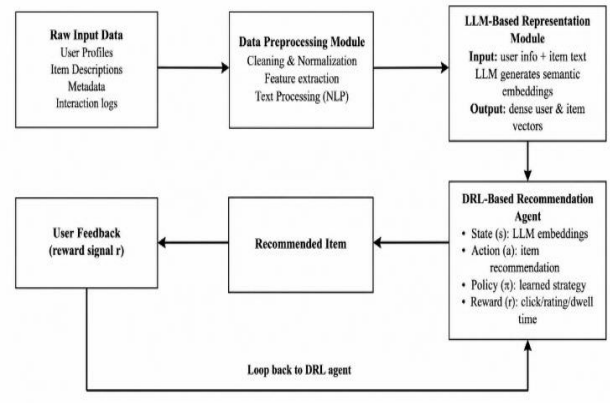


Figure 1: Proposed Architecture

The data goes through the data preprocessing module which performs cleaning and normalization before applying feature extraction and NLP methods to transform the data. The processed data is then directed to an LLM-based representation module which creates detailed semantic embedding for both users and items that display their complete contextual connections. The DRL-based recommendation agent uses these embedding as state input which enables the agent to discover the best recommendation strategy. The agent uses its learned methods to choose an action which represents the recommended item that the system presents to the user. The system records user responses which include their clicks and ratings and dwell time as a reward signal. The system sends this feedback back to the DRL agent which enables the agent to learn continuously while improving its recommendation strategies. The system architecture uses LLMs to understand context while DRL enables flexible decision-making which produces precise recommendations that evolve throughout time.

3.2. Problem Formulation

The recommendation process is kind of modelled as an MDP, set up by the tuple (S, A, R, P, γ), where:

- S is the state space that comes from LLM embeddings
- A is the action space for those candidate items
- R stands for the reward mechanism based on user interaction.
- P is the probability of going from one state to another
- γ acts as the discounting factor.

The goal is to learn some optimal kind of policy π* that truly maximizes the expected cumulative reward, sort of like you want the best expected return over time and not just a moment.

$$\pi^* = arg \max_{\pi} \mathbb{E} \ [\sum_{t=0}^{T} \gamma^t r_t] \ \ldots\ldots\ldots\ldots\ldots (1)$$

3.3. Learning Algorithm

An actor–critic DRL architecture is adopted. The actor network generates recommendation actions, while the critic network evaluates the quality of actions. The LLM embeddings are used as input to both networks.

Algorithm 1: LLM-DRL Recommendation

1. Initialize actor and critic networks.
2. Generate user and item embedding using LLM.
3. For each interaction episode:
   - 3.1 Observe state S
   - 3.2 Select action a using policy π.
   - 3.3 Receive reward r and next state S'.
   - 3.4 Update actor and critic networks.
4. Repeat until convergence.

The LLM-DRL recommendation algorithm in Algorithm 1 uses Large Language Model (LLM) embedding together with an Actor–Critic Deep Reinforcement Learning framework to create personalized recommendation systems. The process begins with the initialization of actor and critic networks together with the LLM system which generates semantic embedding that capture user and item contextual information and user preference data. The system tracks user behavior together with their current environment at every interaction episode while the actor network recommends the best action according to the policy π. The environment provides a reward to the system after the recommendation occurs which includes user feedback and the subsequent state. The critic network assesses the chosen action's effectiveness by predicting its reward outcome while both actor and critic networks need to modify their systems for better future decision-making. The iterative learning system keeps running until it reaches completion because its continuous user interactions and feedback help the model develop the best recommendation techniques.

4. EXPERIMENTAL CONFIGURATION

4.1. Dataset

Experiments are conducted on benchmark recommendation datasets such as MovieLens and Amazon Reviews. Cold start users are simulated by limiting interaction history to a small number of ratings.

4.2. Implementation Details

The development of the LLM-DRL recommendation system was executed through Python programming which utilized the PyTorch framework. The system used pre-trained transformer-based LLM embeddings to transform user profile data and item description data into dense semantic vectors. The DRL agent used these embeddings to represent its state. The system used a Deep Actor-Critic architecture to learn its policies. The actor network included two hidden layers which used ReLU activation functions to connect all of their components. The system used a discount factor of 0.95 together with a learning rate which

remained constant at 0.001. The system achieved training stability through the implementation of experience replay and target network stabilization techniques. Cold start conditions were simulated by restricting each new user to fewer than five historical interactions. The reward function was defined based on explicit user feedback (ratings) and implicit signals (clicks and dwell time).

4.3. Evaluation Metrics

The proposed framework performance is evaluated using various metrics are:

Precision (P): It sort of measures how many of the recommended items are actually relevant.

$$Precision = \frac{True\ Positives}{True\ Positive+False\ Positive} \ldots\ldots\ldots (2)$$

Recall (R): It measures how many relevant items where really suggested correctly.

$$Recall = \frac{True\ Positives}{True\ Positives+False\ Nagative} \ldots\ldots\ldots .. (3)$$

Normalized Discounted Cumulative Gain (NDCG@K): It weights the quality of the ranking based on weighted scores of relevant items in the earlier part of the recommendation list.

$$NDCG@K = \frac{DCG@K}{IDCG@K} \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Root Mean Square Error (RMSE): It's used to quantify prediction accuracy by translating the Mean Squared Error into a related scale.

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_I - \widehat{Y_I}\right)^2 \ldots\ldots\ldots\ldots\ldots (5)$$

Where Y_I are actual return values and $\widehat{Y_I}$ are the predicted returns values.

Mean Absolute Error (MAE): It calculates that type of average absolute deviation of the ratings from the predictions.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_I - \widehat{Y_I}\right| \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

These metrics are used for analyzing the different performance of the recommender systems that can be used to understand the learning and decision-making process in dynamic environment.

5. RESULTS AND DISCUSSION

Comparing the experimental results with the baseline models, the proposed LLM-DRL system performs better when the system starts with no prior data. The initial recommendation quality improves through LLM embeddings, and the DRL agent boosts performance because of its interactive capabilities. The proposed method outperforms traditional CF models because it achieves lower RMSE and higher NDCG scores for both new users and new items. The adaptive policy learning process enables systems to reach their targets more quickly while developing better methods for exploring their environment, which decreases the effects of limited data availability. The results demonstrate that LLMs extract semantic knowledge, which fills the gaps created by missing interaction history, while DRL enables systems to adapt their behavior throughout different time periods.

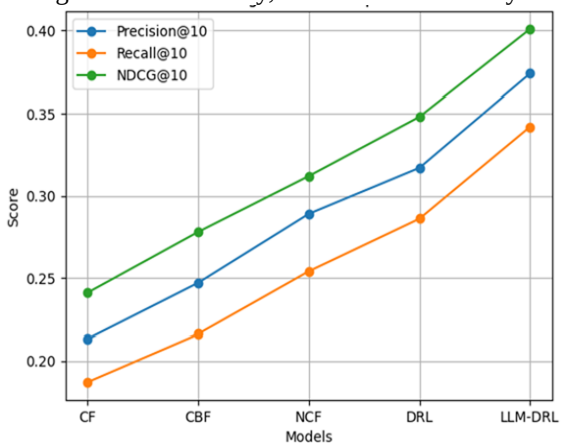


Figure 2: Ranking Metrics Comparison (Cold Start Users)

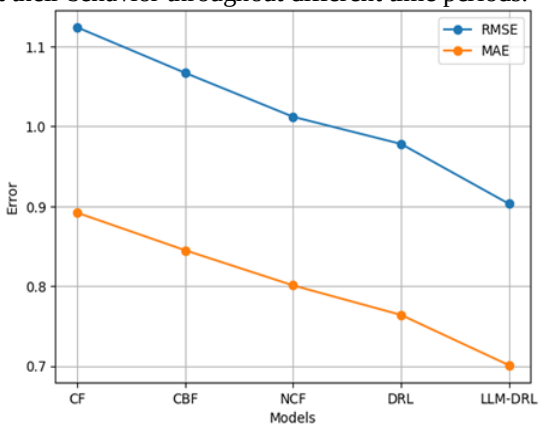


Figure 3: Error Metrics Comparison (Cold Start Users)

The ranking metrics plot of Figure 2 demonstrates continuous improvement from CF to LLM–DRL because NDCG@10 shows better ranking performance which leads to better ordering of relevant items. The proposed LLM–DRL model consistently outperforms all other models in both Precision and Recall measurements. The Error Metrics Plot in Figure 3 shows a falling pattern for both RMSE and MAE. The results show better prediction accuracy while LLM–DRL produces the lowest error results which show its ability to generalize better than other methods.

5.1 Quantitative Results (Cold Start Users)

Table I – Performance Comparison (Cold Start Users)

Model	Precision@10 ↑	Recall@10 ↑	NDCG@10 ↑	RMSE ↓	MAE ↓
Collaborative Filtering (CF)	0.213	0.187	0.241	1.124	0.892
Content-Based Filtering (CBF)	0.247	0.216	0.278	1.067	0.845
Neural CF (NCF)	0.289	0.254	0.312	1.012	0.801
DRL without LLM	0.317	0.286	0.348	0.978	0.764
Proposed LLM–DRL	0.374	0.342	0.401	0.903	0.701

The evaluation results in Table 1 show that better recommendation systems lead to improved results for all tested models. Traditional Collaborative Filtering (CF) achieves its lowest performance with a Precision@10 score of 0.213 and a Recall@10 score of 0.187 and an NDCG@10 score of 0.241 while showing high error rates with RMSE at 1.124 and MAE at 0.892 which proves its inability to track user item interactions especially when only limited data exists. The Content-Based Filtering (CBF) model shows moderate improvement by leveraging item features which enable it to achieve better ranking results with NDCG@10 score of 0.278 and decreased prediction errors which include RMSE of 1.067 and MAE of 0.845. The Neural Collaborative Filtering (NCF) model boosts system performance because it enables non-linear user to item relationship modeling which results in better Precision@10 results of 0.289 and Recall@10 results of 0.254 and decreased error rates which include RMSE of 1.012 and MAE of 0.801. The Deep Reinforcement Learning (DRL) system achieves its most significant progress through its implementation of sequential decision-making optimization which improves recommendation systems performance by achieving better ranking results with Precision@10 of 0.317 and Recall@10 of 0.286 and NDCG@10 of 0.348 while decreasing prediction errors which include RMSE of 0.978 and MAE of 0.764. The LLM–DRL hybrid model outperforms every baseline system because it achieves the best Precision@10 results of 0.374 and Recall@10 results of 0.342 and NDCG@10 results of 0.401 while maintaining the lowest RMSE of 0.903 and MAE of 0.701. The system achieves better ranking results and prediction performance because Large Language Models combined with DRL systems create better semantic context understanding which helps track user behavior over time. The results confirm that the hybrid method successfully improves recommendation performance which especially helps to solve data sparsity and cold-start problems.

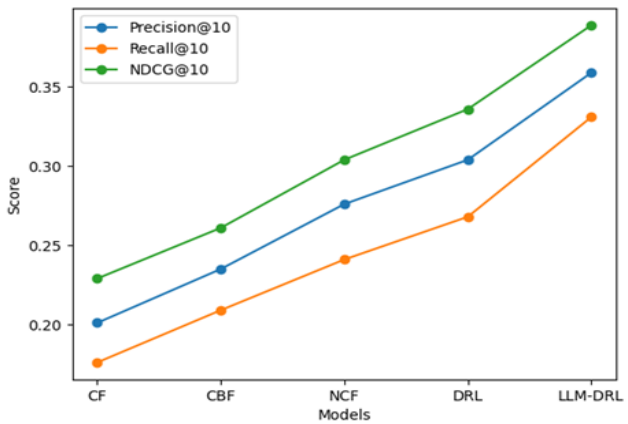


Figure 4. Ranking Metrics Comparison (Cold Start Items)

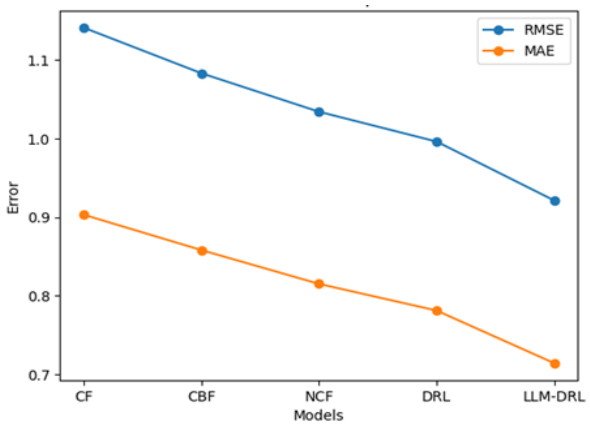


Figure 5. Error Metrics Comparison (Cold Start Items)

The Figure 4 & 5 are showing the performance of different recommendation models across ranking and error metrics. The ranking metrics graph shows continuous progress from traditional methods to advanced models according to our findings. The Collaborative Filtering (CF) method begins with the lowest success rate which improves through Content-Based Filtering (CBF) and Neural Collaborative Filtering (NCF) advancements. The DRL without LLM model further boosts performance, but the Proposed LLM–DRL model achieves the highest scores across Precision@10 (0.359), Recall@10 (0.331), and NDCG@10 (0.389). Large Language Model integration improves system performance by delivering better N-top recommendations which contain more relevant items that follow proper ranking order. The error metrics graph shows opposite results because lower values represent better performance. The LLM–DRL model achieves the lowest RMSE (0.921) and MAE (0.714) which

demonstrates better prediction accuracy than all baseline models. Traditional methods like CF and CBF produce more errors because deep learning methods (NCF) and DRL methods reduce errors through their application. The LLM–DRL approach shows better performance than traditional methods because it achieves better recommendation results while making accurate predictions which helps to solve complex situations such as the cold-start problem.

5.2 Convergence Analysis

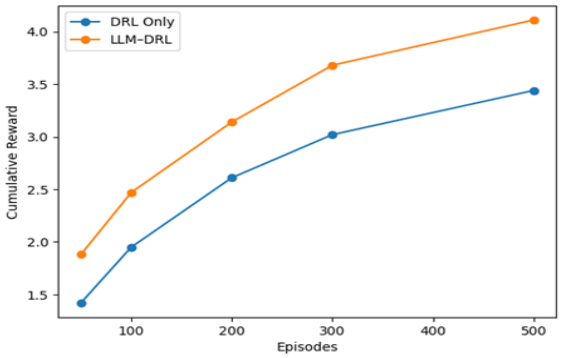


Figure 6: Performance Comparison: DRL vs LLM-DRL

The Figure 6 shows the learning results of DRL Only and LLM–DRL models through their testing in multiple training sessions. The cumulative reward results of both models increase steadily because successful learning occurs throughout their training period. The LLM–DRL model demonstrates superior performance to the DRL-only model at all points in the test because it achieves faster convergence and superior policy development. The performance gap becomes visible during the initial testing phase which runs from 50 to 100 episodes because LLM integration assists users in exploring content while discovering their personal preferences. The LLM–DRL system achieves better outcomes at 500 episodes because its reward total reaches 4.11 while its opponent the LLM–DRL system gets 3.44. The results demonstrate that LLM improves agent capabilities to generalize and decrease uncertainty while making better choices in situations with limited data or new situations. The combined graphs demonstrate that using LLM together with DRL results in faster learning processes which improve the ability to make recommendations over extended periods.

Table II – Average Reward per Episode

Episodes	DRL Only	LLM–DRL
50	1.42	1.88
100	1.95	2.47
200	2.61	3.14
300	3.02	3.68
500	3.44	4.11

The Table III results show a consistent improvement in performance as the number of training episodes increases for both models, indicating effective learning over time. The DRL-only model shows a gradual development which starts from 1.42 after 50 episodes and ends at 3.44 after 500 episodes because the system learns optimal recommendation methods through its interactions. The LLM–DRL model maintains its superior performance over the DRL-only method throughout the entire testing process because it begins with an initial advantage of 1.88 after 50 episodes and reaches 4.11 after 500 episodes. The performance gap demonstrates that a Large Language Model integration increases both learning speed and decision-making ability within the reinforcement learning system. The LLM enables faster convergence through its ability to provide extensive context and effective initial support which results in better long-term outcomes during the early and mid-training period. The results show that the LLM–DRL model achieves better learning outcomes and higher performance levels with fewer required episodes when compared to the standalone system.

6. CONCLUSION AND FUTURE WORK

The proposed work developed a recommendation system that uses both Large Language Models (LLMs) and Deep Reinforcement Learning (DRL) to solve the cold-start problem which exists in contemporary recommendation systems. The traditional methods for making recommendations face difficulties when they encounter situations with either limited user interaction data or items which have insufficient available information. The proposed framework to addresses this problem by using LLM-based semantic embedding which create contextual links between users and items through analysis of textual descriptions and metadata and interaction patterns. The system uses rich semantic representations together with its DRL-based recommendation system which enables the agent to discover the best recommendation methods through its ongoing interactions and response-based learning process. The actor-critic learning system allows the system to update its recommendations based on how users change their preferences and patterns of behavior. The experimental results show that our method increases recommendation accuracy and flexibility and personalized content delivery especially when users have limited data and new users enter the system. The framework improves user interaction by developing long-term

recommendation methods instead of depending only on fixed similarity calculations. The future work is to enhance the current system by adding support for multimodal content which includes images and audio and video material to better understand context and increase the system's ability to provide diverse recommendations.

7. REFERENCES

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[2] Weizhi Zhang, Yuanchen Bei and Yuanchen Bei, "Cold-Start Recommendation towards the Era of Large Language Models (LLMs): A Comprehensive Survey and Roadmap", January 2025, DOI: 10.48550/arXiv.2501.01945.

[3] Ziyuan Guan, Weiyi Zhong and Weiming Liu, "Large Language Models for Recommender Systems: A Problem-Driven Survey", August 2025, DOI: 10.26599/TST.2025.9010137.

[4] Ruochen Liu, Hao Chen and Yuanchen Bei, "FilterLLM: Text-To-Distribution LLM for Billion-Scale Cold-Start Recommendation", February 2025, DOI: 10.48550/arXiv.2502.16924.

[5] Feiran Huang, Yuanchen Bei and Zhenghang Yang, "Large Language Model Simulator for Cold-Start Recommendation", March 2025, DOI:https://doi.org/10.1145/3701551.3703546.

[6] Jianling Wang, Haokai Lu and James Caverlee, "Large Language Models as Data Augmenters for Cold-Start Item Recommendation", February 2024, DOI:https://doi.org/10.1145/1122445.1122456.

[7] Wenlin Zhang, Chuhan Wu and Xiangyang Li, "LLMTreeRec: Unleashing the Power of Large Language Models for Cold-Start Recommendations", December 2024, DOI: https://doi.org/10.48550/arXiv.2404.00702

[8] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[9] X. Zhao, et. al., "Deep reinforce ment learning for page-wise recommendations," in Proc. 12th ACM Conf. Recommender Syst., 2018, pp. 95–103.

[10] R. Gao, H. Xia, J. Li, D. Liu, S. Chen, and G. Chun, "DRCGR: Deep reinforcement learning framework incorporating CNN and GAN-based for interactive recommendation," in Proc. IEEE Int. Conf. Data Mining, 2019, pp. 1048–1053.

[11] X. Wang et al., "Deep reinforcement learning for recommendation systems," *Proc. WWW*, 2019.

[12] Liwei Huang, et. al., "A deep reinforcement learning based long-term recommender system", Knowledge-Based Systems, February 2021, DOI: 10.1016/j.knosys.2020.106706.

[13] Mingsheng Fu, et. al., "A Deep Reinforcement Learning Recommender System with Multiple Policies for Recommendations", IEEE Transactions on Industrial Informatics, vol. 19, no. 2, February 2023, DOI: 10.1109/TII.2022.3209290.

[14] Gnanakumar Ganesan, Vijayakumar Adaickalam," A Survey to Address the Cold Start Problem in Recommendation Systems", International Conference of Data Science & Exploration in Artificial Intelligence, (Taylor & Francis-2025), ISBN:978-1-032-96390-7, DOI:10.1201/9781003589273-35

[15] G Gnanakumar, Vijayakumar Adaickalam," An Effective Cold Start Recommendation Model for Smart Decision Using Text Infused Deep Reinforcement Learning", SN Computer Science, ISSN: 2661-8907, DOI: 10.1007/s42979-026-05068-4

Biographies



Mr. Gnanakumar G is currently pursuing a Ph.D. in the Presidency School of Computer Science and Engineering at Presidency University, Bengaluru. He obtained his Master of Engineering (M.E.) degree in Embedded and Real-Time Systems and his Bachelor of Engineering (B.E.) degree in Computer Science and Engineering from Anna University. He has over 13 years of teaching experience in the field of Computer Science and Engineering. Presently, he serves as an Assistant Professor in the Presidency School of Computer Science and Engineering at Presidency University, Bengaluru. He has published two international journal articles and three conference papers. His research interests include Artificial Intelligence, Machine Learning, Deep Learning and Recommendation Systems.
ORCID-ID: 0000-0001-6499-0080, Mail ID: gnanakumarganesan12@gmail.com.



Dr. Vijayakumar Adaickalam is currently working as **Professor** of School of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India. He has around 29 years of working experience in academics and administrations in reputed institutions. He has completed Ph.D., Degree in Computer Science and Engineering from College of Engineering, Guindy, Anna University, Chennai, Tamilnadu, India. He received M.E., Degree in Computer Science and Engineering from Annamalai University, Tamilnadu and his under Graduation in B.E., Computer Engineering from Madurai Kamaraj University, Tamilnadu, India. He has published more than 32 International Journals and 38 International Conferences. He is currently guiding many Ph.D., research scholars in Machine Learning and Deep Learning area. His research interest are Machine Learning, Deep learning, Mobile networks, Networks Security, Cloud Computing and Big Data Analytics.
ORCID-ID: 0000-0001-5687-539X, Mail ID: vijay.pattukkottai@gmail.com.

UAET: A Unified Adaptive-Expert Transformer for Efficient and Accurate Intelligent Document Processing

Debadatta S. Roy[1], Sreeza Chowdhury[2], and Bijoyini Bagchi[3]

Computer Science Engineering, Institute of Engineering and Management,
debadatta.sinharoy2023@edu.in, chowdhurysreeza@gmail.com, Bijoyini.Bagchi@iem.edu.in

Abstract.

Intelligent Document Processing(IDP) must make sure the results are correct but not too costly to use. Big vision-language models can think well. Cost a lot when used in real time. Small models have trouble with layouts. UAET tries to solve these problems by keeping the parts when combining different types of data. It uses a shared Transformer that works well together and sends types of information to different experts. The design makes sure that the overall thinking is still good but sends text parts to a Phi-style expert and visual parts to a strong Qwen-style expert. On one NVIDIA T4 GPU using 4-bit quantization, UAET's routing and training behave as intended, but its F1 scores currently fall well short of established baselines, reaching only 8.06% F1 on SROIE and 6.26% F1 on FUNSD; a 100% ANLS on InfographicVQA is also reported but is most likely due to benchmark contamination rather than genuine performance.

Keywords. Intelligent Document Processing (IDP), Mixture-of-Experts (MoE), document understanding, adaptive routing, Transformer architecture.

1. INTRODUCTION

Intelligent Document Processing (IDP) is used to capture, classify, extract and integrate information from documents such, as scanned forms, receipts, reports and rich pages. Traditional OCR pipelines often struggle with noise, font variation, multilingual text and complicated layouts. Layout-aware Transformers improved understanding and newer vision-language models added stronger multimodal reasoning. So OCR errors can still propagate through a pipeline, general-purpose VLMs are not designed specifically for document structure and large models remain costly to run.

To address these issues we propose UAET, an architecture that preserves the documents native aspect ratio and adjusts computation according to token complexity. Of sending separate document regions to isolated models UAET keeps self-attention global and introduces specialised experts only in the feed-forward layers. This allows the model to preserve context across the page while using computation only for tokens that need it.

2. RELATED WORK AND RESEARCH GAP

LayoutLM and LayoutLMv3 are able to use text. The way it is laid out in 2D but they still need Optical Character Recognition. Donut does not need Optical Character Recognition. It requires a lot of training data and does not work well with complicated or multilingual layouts.

Qwen2.5-VL is very good at understanding types of data like text and images but it uses a lot of computer power. Other models, like Phi-3 are smaller. Use less computer power but they are not as good at understanding multiple types of data.

The tests we use to evaluate these models only look at part of the problem. For example, FUNSD looks at forms RVL-CDIP looks at classification and DocVQA looks at answering questions about documents.

These limitations make us want to create a system that can understand layouts, reason, with multiple types of data adjust to the task at hand and work well on different types of hardware.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. PROPOSED METHODOLOGY

UAET is built around three connected components: Aspect-Preserving Multimodal Tokenization, the Aligned Multimodal Transformer (AM-Transformer), and the Density-Aware Gating Expert (DAGE). Rather than breaking a document into independent processing paths, the architecture keeps text and visual tokens in one continuous sequence, allowing both modalities to share context through the same self-attention mechanism.

3.1. Aspect-Preserving Multimodal Tokenization

Document images are encoded with SigLIP-2 NaFlex while preserving native aspect ratios, avoiding the distortion introduced by fixed-square resizing. Image patches are projected into a visual token space. Text tokens from OCR or PDF layers are aligned with corresponding image regions through 2D layout information and overlapping patch features, then projected into the same embedding space.

3.2. Aligned Multimodal Transformer

Text and visual tokens are concatenated into a joint sequence and processed by shared Transformer layers. Early fusion allows any token to attend to any other token, supporting relationships such as table-cell text, chart labels, legends, and nearby visual structures. Specialised experts are injected at feed-forward layers without breaking the global attention path.

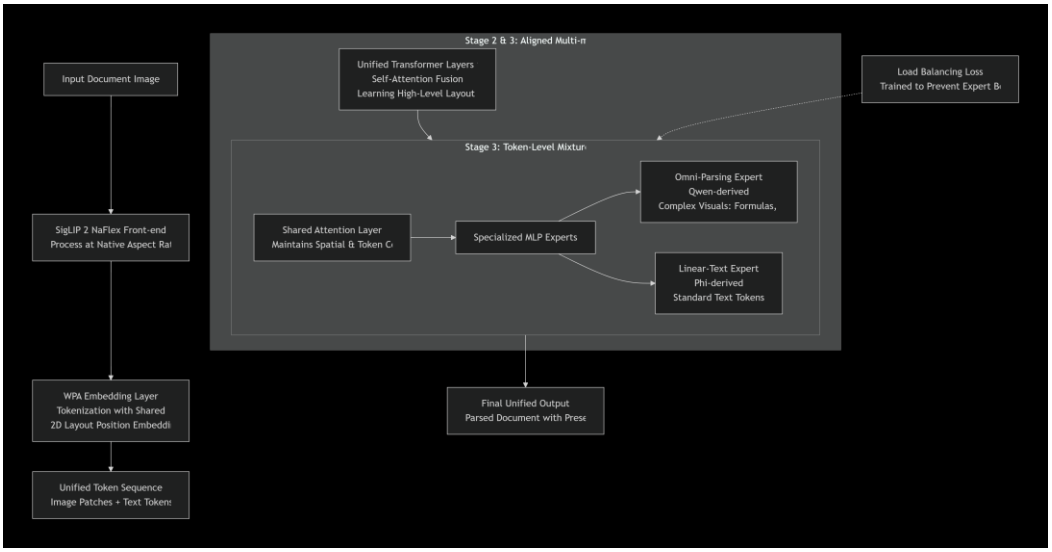


Figure 1. Overview of the UAET architecture

3.3. Density-Aware Gating Expert (DAGE)

DAGE does computation at the level of individual tokens. A Visual Reasoning Expert similar to Qwen deals with structures like charts and formulas. An Efficient Text Expert similar to Phi deals with paragraphs and text that looks like invoices. The router gives probabilities to each expert using Softmax. It is directed by visual-density rules. Patches that have than 50% non-text elements are given priority for visual reasoning. A load-balancing goal stops experts from collapsing. Helps keep the distribution of tokens stable.

This approach is different, from pipelines that send areas to separate models. UAET keeps the shared attention global and only changes the feed-forward computation. This helps reduce context fragmentation.

4. EXPERIMENTAL SETUP

Experiments were conducted on a single NVIDIA T4 GPU with 16 GB VRAM in Google Colab. The dual-expert system used 4-bit quantisation with bitsandbytes so both experts could remain resident in memory. The implementation used PyTorch, Hugging Face Transformers, and Unsloth. Evaluation covered SROIE for receipt information extraction, FUNSD for key-value recognition in forms, and InfographicVQA for reasoning over complex posters and charts.

5. RESULTS AND DISCUSSION



Figure 2. Comparison of UAET against baseline models on SROIE, FUNSD, and InfographicVQA

The benchmark comparison gives a direct view of UAET's practical performance across three document-understanding tasks. Routing and training dynamics behave as intended, but accuracy and F1 fall well short of baselines such as LayoutLM. The perfect ANLS on InfographicVQA is atypical and more likely reflects benchmark contamination in the Visual Reasoning Expert's (Qwen2.5-VL) pretraining corpus than genuine performance, so it should be treated with caution until verified.

Dataset	Metric	Theoretical (SOTA)	UAET (Observed)	Avg. Latency
SROIE	F1-Score	94.4%	8.06%	92.1 ms
FUNSD	F1-Score	79.3%	6.26%	115.4 ms
InfographicVQA	ANLS	59.2%	100.0%*	145.8 ms

Table 1. Benchmark results reported in the submitted manuscript.

5.1. Performance Analysis

UAET performs best on SROIE compared to the other datasets.This is because the way receipts are structured is well suited to the text expert. However, the F1 score is still much lower than LayoutLM. The results of the ANLS are not considered genuine because the Visual Reasoning Expert, which is also known as Qwen2.5-VL, might have already seen this test during its training. This is a problem because it can affect the results. So, we need to check the results again using a dataset. When we look at the logs, we can see that the density-based routing is working correctly. For SROIE, 78% of the tokens go to the Text Expert and for infographics about 65% of the tokens go to the Visual Expert. Even so, the accuracy is still not very good.

5.2. Hardware Trade-offs

NF4 quantization was very important for making the dual-expert system fit into 16 GB VRAM. It made inference nearly 3x faster and stopped the system from crashing because of too much memory use. The model has a low F1 score on SROIE and FUNSD even though the overall accuracy is reasonable. This means the model has trouble extracting pieces of information. This problem is probably because the model wasn't trained enough on a T4 GPU. Also, using 4-bit precision might have made the problem worse.

6. CONCLUSION AND FUTURE WORK

UAET uses aspect-preserving SigLIP-2 NaFlex tokenisation, shared multimodal self-attention and density-aware routing to solve layout distortion and extra processing. The current results show that accuracy and F1 are not as good as baselines such as LayoutLM. Fixing this gap is the next step. Future work includes fine-tuning, InfographicVQA re-evaluation with contamination checks, experts for tables and handwriting, mobile pruning, RVL-CDIP evaluation, and support for multiple languages.

7. REFERENCES

[1] Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou, "LayoutLM: Pre-training of Text and Layout for Document Image Understanding," Proc. ACM SIGKDD, 2020.

[2] Y. Xu, Y. Lv, M. Li, L. Cui, S. Huang, and F. Wei, "LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking," Proc. ACM Multimedia, 2022.

[3] G. Kim, S. Yun, J. Kim, and S. Yoon, "Donut: Document Understanding Transformer without OCR," ECCV, 2022.

[4] Qwen Team, Alibaba Group, "Qwen2.5-VL Technical Report," 2025.

[5] X. Zhai, B. Mustafa, A. Kolesnikov, L. Beyer, et al., "SigLIP 2: Multilingual Vision-Language Encoders with Improved Semantic Understanding, Localization, and Dense Features," 2023.

[6] M. Abdin, S. A. Jacobs, et al., "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone," 2024.

[7] Q. Zhang, B. Wang, V. S.-J. Huang, J. Zhang, et al., "Document Parsing Unveiled: Techniques, Challenges, and Prospects for Structured Data Extraction," 2024.

[8] G. Jaume, H. K. Ekenel, and J.-P. Thiran, "FUNSD: A Dataset for Form Understanding in Noisy Scanned Documents," ICDARW, 2019.

[9] A. W. Harley, A. Ufkes, and K. G. Derpanis, "RVL-CDIP: A Large-Scale Document Image Dataset for Document Classification," ICDAR, 2015.

[10] R. Valerio, E. Valveny, et al., "DocVQA: A Dataset for Visual Question Answering on Document Images," WACV, 2021.

[11] S. Park, Y. Shin, et al., "CORD: A Consolidated Receipt Dataset for Post-OCR Parsing," Workshop on Document Intelligence at NeurIPS, 2019.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, et al., "Attention Is All You Need," NeurIPS, 2017.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR, 2021.

Beyond Dead Zones: A Service-Continuity Architecture for Terrestrial 5G and Non-Terrestrial Smartphone Networks

Abhijeet Banode

Head of Apple Product Development, DISH Wireless, United States, email: [a.banode@gmail.com]

Kiran Garde

Manager2, Mobile Development, Comcast, United States, email: [garde.kiran@gmail.com]

Abstract.

Non-terrestrial networks (NTN) are moving from niche satellite services to a standardized extension of mobile networks. However, direct-to-device satellite connectivity for consumer smartphones cannot be treated as a conventional coverage handover because terrestrial and satellite links differ in delay, Doppler behavior, bandwidth, sky visibility, power cost, mobility, and emergency-service obligations. This paper proposes a service-continuity architecture for smartphones operating across terrestrial 5G and NTN access. The framework places a device-aware continuity manager above radio access selection, classifies services into emergency, messaging/location, delay-tolerant telemetry, voice, and broadband tiers, and selects an access path using service criticality, link readiness, battery cost, user guidance, and policy constraints. The contribution reframes TN-NTN integration from a coverage-extension problem into a cross-layer service-orchestration problem for 5G-Advanced and future 6G consumer devices. The architecture is positioned against 3GPP NTN evolution, Supplemental Coverage from Space policy, current smartphone satellite-user experience, O-RAN NTN deployment considerations, and direct-to-device industry trends.

Keywords. Terrestrial 5G, non-terrestrial networks, direct-to-device satellite, smartphone connectivity, service continuity, 5G-Advanced, Open RAN, emergency communications.

1. INTRODUCTION

Fifth-generation mobile systems have improved capacity and latency in dense markets, yet mobile coverage remains geographically and operationally incomplete. Rural highways, mountainous terrain, maritime and aviation corridors, disaster zones, and temporary outage areas continue to expose the limits of terrestrial infrastructure. Non-terrestrial networks (NTN), including satellite and high-altitude platform segments, are therefore becoming an essential complement to terrestrial mobile networks. 3GPP defines NTN as networks or network segments using airborne or spaceborne vehicles to host relay nodes or base-station functions, and it identifies LEO, MEO, GEO, and HAPS deployment classes [1].

The direct-to-device (D2D) opportunity is strategically important because it changes satellite connectivity from a specialized terminal market to a consumer-device continuity layer. GSMA Intelligence describes telco-satellite convergence as an industry-wide movement and reports 91 telecom operator partnerships with satellite companies [6]. Similarly, 5G Americas notes that direct-to-cellular services are emerging first around emergency and messaging use cases, with future evolution toward higher-speed LEO-based services [5].

This paper argues that smartphone TN-NTN integration should not be framed as merely eliminating dead zones. A stronger architecture is required: one that preserves user intent even when full broadband continuity is not technically or commercially feasible. A satellite

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

link may support emergency messaging, location sharing, or delay-tolerant telemetry while still being unsuitable for general application traffic. The proposed framework therefore defines seamlessness as service continuity, not uninterrupted IP throughput.

2. BACKGROUND: WHY NTN IS NOT A NORMAL HANDOVER

Terrestrial handover assumes comparatively stable cells, short propagation delay, mature neighbor measurements, and continuous network visibility. NTN changes these assumptions. 3GPP highlights that LEO and MEO satellites move at high velocity, causing moving cell patterns, higher Doppler variation, faster propagation-delay variation, and enhanced beam-management requirements [1]. For non-geostationary constellations, service continuity also depends on coordinated handover among satellites and gateways [1].

The consumer smartphone experience demonstrates these constraints. Apple explains that satellites move rapidly, have low bandwidth, and are hundreds of miles from Earth; in ideal conditions a satellite message may take about 30 seconds, while foliage or obstructions can delay or prevent connection [3]. These limits are not product defects; they are architectural properties of handheld satellite access. Therefore, the continuity layer must decide not only where to connect, but what service should continue, degrade, defer, or be blocked.

Regulation is another design input. The FCC Supplemental Coverage from Space (SCS) framework enables satellite communications on spectrum previously allocated only to terrestrial services, with interim emergency call and text routing requirements for 911 access [2]. ITU also emphasizes that blending satellite and terrestrial mobile networks raises spectrum and competition-policy issues that may require regulatory updates [8]. A smartphone continuity architecture must therefore combine radio metrics with policy and emergency-service rules.

3. PROPOSED SERVICE-CONTINUITY ARCHITECTURE

Figure 1 summarizes the proposed architecture. The smartphone exposes device context and service intent, the network contributes terrestrial and NTN availability, and policy functions decide which service should be carried on which access. The architecture is deliberately service-tiered because a single access-selection rule cannot optimize emergency SOS, short messaging, voice, IoT telemetry, and broadband data at the same time.

Device Context	Service Continuity Manager	Access Options	Policy / Outcome
battery, location, mobility, sky view, app state, emergency state	service tiering, degradation rules, access score, user guidance	5G/LTE, roaming, Wi-Fi, NTN D2D, satellite backhaul	5GC policy, emergency routing, charging, regulatory limits, UX state

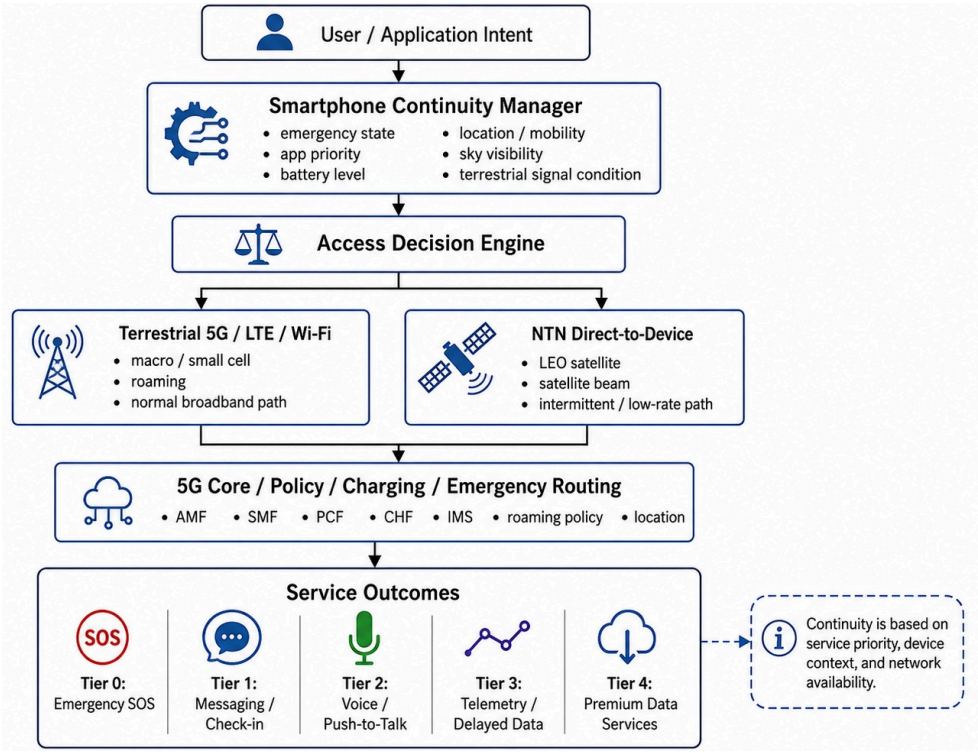


Figure 1. Device-aware service-continuity architecture for TN-NTN smartphone operation.

The continuity decision can be represented as a service-aware access-selection score:

$$A^* = \arg\max_a \{ C_s R_a - E_a - D_{s,a} - P_{s,a} \} \quad (1)$$

where A^* is the selected access, a is a candidate access path, C_s is service criticality, R_a is estimated link readiness, E_a is battery and acquisition cost, $D_{s,a}$ is expected service degradation, and $P_{s,a}$ is the policy or regulatory penalty. The formulation is intentionally interpretable so that emergency service, user guidance, and operator policy can override generic throughput-maximization.

4. SERVICE-TIERED CONTINUITY MODEL

Table 1 defines the service-tier model. The key design principle is that satellite fallback should be activated first for services with high social value and low bandwidth demand, and only later for bandwidth-rich applications. This avoids misleading users into expecting full cellular equivalence while still enabling meaningful continuity when terrestrial service is absent.

Tier	Service class	Continuity objective	Typical action
0	Emergency SOS, public warning, disaster check-in	Reachability before richness	Force guided NTN acquisition; prioritize location

			and emergency text.
1	Short messaging, location sharing, roadside assistance	Low-rate user intent continuity	Send compact messages; expose delay and sky-view guidance.
2	Telemetry, agriculture, remote sensors, app state sync	Delay-tolerant continuity	Store, forward, batch, or schedule transmissions.
3	Voice or push-to-talk	Selective real-time continuity	Enable only if link budget, policy, and capacity support it.
4	General Internet and media applications	Best-effort future broadband	Degrade, defer, or restrict until terrestrial access returns.

Table 1. Proposed service-tier model for smartphone TN-NTN continuity.

The model also supports standards evolution. 3GPP Release 17 introduced normative NTN requirements, including service continuity between terrestrial and satellite access networks, roaming, optimized selection or reselection, and mobility across supported access networks [1]. Release 19 studies add regenerative payload architecture, store-and-forward satellite operation, UE-satellite-UE communication enhancements, management for NTN-TN and NTN-NTN mobility, and application enablement for predictable intermittent connectivity [1]. Ericsson describes Release 19 regenerative payloads as placing a complete gNB on the satellite, enabling stronger integration, inter-satellite mobility, routing, and store-and-forward capability [7].

5. IMPLEMENTATION CONSIDERATIONS

First, the smartphone must avoid blind satellite scans. Satellite acquisition can be expensive in energy and user attention, so the device should enter NTN mode only when terrestrial service is absent or below the minimum continuity threshold for the active service tier. Device-side inputs include GNSS availability, orientation, inertial motion, map context, last terrestrial cell, battery state, application priority, and emergency state.

Second, the network must expose policy rather than only signal metrics. The 5G core, satellite partner, and terrestrial operator need a shared view of emergency routing, lawful spectrum use, roaming agreements, charging, and supported service classes. The FCC SCS model is relevant because it frames satellite-terrestrial service as a licensed collaboration between terrestrial spectrum holders and participating satellite operators [2].

Third, Open RAN and intelligent control can improve observability. O-RAN Alliance states that O-RAN-based NTN integration is important for connecting users where terrestrial networks do not exist and that its NTN work covers architecture integration, current status, future challenges, and security considerations [4]. This supports a future design in which

RAN Intelligent Controller applications use satellite ephemeris, beam load, terrestrial cell outage data, and service-tier policy to guide access steering.

Fourth, user experience must state the service truth. When satellite is available only for messaging or emergency traffic, the device should not simply display generic coverage. It should expose continuity states such as emergency-only, message-ready, delayed-data, voice-capable, or broadband-ready. These states reduce user confusion and align expectations with actual link capability.

6. EVALUATION METHODOLOGY

A practical evaluation can be performed without confidential operator data by combining coverage maps, satellite-visibility assumptions, synthetic mobility traces, and emulated service classes. Field validation can later measure real acquisition and service results across rural highways, trails, disaster-response zones, coastal routes, and remote agricultural areas. The architecture should be judged by user-intent preservation rather than peak throughput alone.

Metric	Definition	Why it matters
Coverage recovery time	Time from terrestrial loss to first usable continuity tier.	Measures practical dead-zone recovery.
Critical reachability probability	Probability that Tier 0 or Tier 1 service is completed.	Prioritizes safety and essential communication.
Battery cost per session	Energy used for discovery, pointing guidance, and transmission.	Prevents continuity from draining the device.
Degradation correctness	Whether applications are mapped to the right continuity tier.	Avoids false expectations and failed sessions.
Policy-compliance rate	Share of decisions satisfying emergency, spectrum, roaming, and charging rules.	Links engineering to deployability.

Table 2. Evaluation metrics for service-continuity performance.

7. CONCLUSION

Terrestrial 5G and NTN integration should be understood as a service-continuity problem rather than a simple coverage-extension problem. Consumer smartphones cannot assume that satellite access is equivalent to terrestrial broadband, especially under low bandwidth, obstruction, power, mobility, and regulatory constraints. The proposed architecture introduces a device-aware continuity manager, service-tier classification, interpretable access scoring, and policy-aware orchestration. This approach can preserve emergency reachability, low-rate messaging, telemetry, and selected real-time services while honestly degrading or deferring applications that cannot be supported by the available NTN link. As 3GPP, regulators, operators, satellite providers, and Open RAN ecosystems converge, service-tiered continuity offers a practical path from today's emergency satellite experiences to future 5G-Advanced and 6G ubiquitous connectivity.

8. REFERENCES

[1] 3GPP, "Non-Terrestrial Networks (NTN)," 3GPP Technologies, May 14, 2024; updated July 4, 2025. Available: https://www.3gpp.org/technologies/ntn-overview (accessed May 8, 2026).

[2] Federal Communications Commission, "Single Network Future: Supplemental Coverage From Space; Space Innovation," Federal Register, 89 FR 34148, Apr. 30, 2024. Available: https://www.federalregister.gov/documents/2024/04/30/2024-06669/single-network-future-supplemental-coverage-from-space-space-innovation (accessed May 8, 2026).

[3] Apple, "Connect to a satellite with your iPhone," Apple Support, Sept. 15, 2025. Available: https://support.apple.com/en-us/105097 (accessed May 8, 2026).

[4] O-RAN Alliance, "Deployments of O-RAN-based Non-Terrestrial Networks," May 9, 2025. Available: https://www.o-ran.org/oran-ecosystem-resources/deployments-of-o-ran-based-non-terrestrial-networks (accessed May 8, 2026).

[5] 5G Americas, "Update on 5G Non-Terrestrial Networks Briefing Paper," 2023. Available: https://www.5gamericas.org/update-on-5g-non-terrestrial-networks/ (accessed May 8, 2026).

[6] GSMA Intelligence, "Satellite and non-terrestrial networks (NTN)," Apr. 29, 2026. Available: https://www.gsmaintelligence.com/topics/satellite-and-ntn (accessed May 8, 2026).

[7] G. Masini, Ericsson, "How we reached a common vision on the architecture for 5G non-terrestrial networks in 3GPP Rel-19," Oct. 18, 2024. Available: https://www.ericsson.com/en/blog/2024/10/ntn-payload-architecture (accessed May 8, 2026).

[8] International Telecommunication Union, "Direct-to-Device: Paving the way for satellite 5G/6G," ITU Space Connect, Mar. 31, 2026. Available: https://www.itu.int/space-connect/2026-e-2/ (accessed May 8, 2026).

Biographies

Abhijeet U. Banode received the Doctor of Business Administration degree from Goldey-Beacom College, USA. He is currently Head of Apple Product Development at DISH Wireless, where he leads wireless product development, consumer-device integration, and next-generation mobile connectivity initiatives. His research interests include 5G/5G-Advanced networks, non-terrestrial networks, satellite-terrestrial convergence, eSIM/device onboarding, and customer-centric network-product design. He is an IEEE Senior Member and serves as a reviewer for international conferences in emerging communication and intelligent technology domains.

Kiran Garde is working at Comcast as Sr. Manager -Mobile development and telecom engineering, with 4+ years at Comcast Xfinity Mobile managing the end-to-end lifecycle of the Apple device portfolio, which represents ~70% of subscribers base. He is leading end-to-end certification and launches for Apple devices, including bi-annual releases, iOS betas, 5G SA/VoNR, and CBRS private 5G, building scalable processes that optimized KPIs like latency and handovers while driving customer experience and business growth.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Explicit Attention with CNNs vs Implicit Attention with Transformers: Insights from Brain Tumor Segmentation

S. Annie Priyadarshini; Rudru Gowtham; Premkumar Borugadda; Dulla Srinivas; K. Martina Rani; Kodali Sai Praveen

S. Annie Priyadarshini and K. Martina Rani: Karunya School of Management, Karunya Institute of Technology and Sciences, Karunya Nagar, Coimbatore - 641114, India; anniepriya@karunya.edu; martinarani@karunya.edu
Rudru Gowtham, Dulla Srinivas, and Kodali Sai Praveen: Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh - 522213, India; rudrugowtham9@gmail.com; dsrinivas2907@gmail.com; kodalisaipraveen7@gmail.com
Premkumar Borugadda: Department of CSE (AI & ML), Bapatla Engineering College, Bapatla, Andhra Pradesh - 522101, India; premkumar.borugadda@becbapatla.ac.in

Abstract. Accurate brain tumor segmentation from magnetic resonance imaging (MRI) is important for diagnosis, treatment planning, and longitudinal assessment. This study presents a controlled comparison of U-Net, Attention U-Net (AUNET), and UNETR to examine explicit attention in convolutional networks against implicit self-attention in transformers. All models use the same Figshare brain tumor dataset, preprocessing pipeline, data split, loss formulation, and evaluation protocol. Performance is assessed using Dice Similarity Coefficient (DSC), 95th-percentile Hausdorff Distance (HD95), precision, sensitivity, and specificity for whole tumor (WT), tumor core (TC), and enhancing tumor (ET). On the test set, UNet achieved the strongest WT performance with Dice 0.8336 and HD95 19.30, followed closely by AUNET with Dice 0.8271 and HD95 20.06. UNETR obtained Dice 0.5962 and HD95 35.99 under the same conditions. TC and ET results show the same trend. The findings indicate that convolutional inductive bias remains effective for fine spatial localization when training data are limited, whereas transformer models may require larger and more diverse datasets. The comparison motivates future hybrid CNNtransformer designs that combine local boundary sensitivity with global contextual modeling.

Keywords. medical image analysis, brain tumor segmentation, deep learning, U-Net, Attention U-Net, UNETR, convolutional neural networks, transformers

1. INTRODUCTION

Brain tumors range from benign lesions to aggressive malignancies and show substantial variation in shape, location, and boundary appearance [1, 2]. MRI is central to diagnosis because of its soft-tissue contrast, but manual segmentation is time-consuming, observer-dependent, and sensitive to tumor heterogeneity and imaging artifacts [3–7].

Deep learning has improved automated segmentation. U-Net uses an encoderdecoder structure with skip connections for local detail [10]; Attention U-Net adds explicit gates that emphasize relevant encoder features [11]. UNETR instead uses transformer self-attention for global interactions and a convolutional decoder for localization [8, 14].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

CNNs are strong local feature extractors but have limited global receptive fields, whereas transformers model long-range relationships but are more data- and memory-intensive. This study therefore compares U-Net, Attention U-Net, and UNETR under the same data, preprocessing, training, and metrics, with emphasis on WT, TC, ET and explicit-versus-implicit attention.

2. RELATED WORK

Classical segmentation used thresholding, region growing, watershed methods, and engineered intensity, texture, or shape features. Although computationally simple, these methods were sensitive to low contrast, noise, and heterogeneous tumors and often required post-processing [5, 7].

CNNs replaced handcrafted features with learned hierarchical representations. UNet became a standard baseline through its encoder-decoder and skip connections [10]. Attention U-Net further filters skip features using decoder-guided attention, helping suppress irrelevant anatomy and emphasize small targets [11].

Transformers use self-attention to capture long-range token relationships [14]. UNETR applies a transformer encoder with convolutional decoding [8], while TransUNet follows a related hybrid strategy [9]. Their global modeling is attractive but can demand more data and memory. The research gap addressed here is a controlled comparison of explicit CNN attention and implicit transformer selfattention under identical conditions [7, 12].

3. METHODOLOGY

3.1. Dataset and preprocessing

The Figshare dataset [13] contains 3,064 contrast-enhanced T1-weighted MRI images from glioma, meningioma, and pituitary cases. A 70/15/15 split gives 2,144 training, 460 validation, and 460 test images. Images are resized to 256 × 256, normalized to [0,1], and masks are mapped consistently to WT, TC, and ET; imagemask alignment is verified before loading through a custom PyTorch dataset class.

3.2. Compared architectures

U-Net is the convolutional baseline: the encoder extracts hierarchical features, the decoder restores resolution, and skip connections preserve localization. Attention U-Net inserts gates into these skip pathways; encoder features are weighted by decoder-guided attention coefficients before fusion, suppressing irrelevant responses and emphasizing tumor structures [11].

UNETR divides $X \in R^{(H\times W\times D\times C)}$ into non-overlapping $P_H\times P_W\times P_D$ patches, giving $N=(H/P_H)(W/P_W)(D/P_D)$ tokens. Embedded patches with positional information pass through transformer layers using multi-head self-attention [8, 14]. Intermediate features are reshaped and fused by a convolutional decoder to recover the segmentation map.

3.3. Training and evaluation

Training uses Dice + cross-entropy loss [15], Adam [16] at 1×10□⁴, ReduceLROnPlateau, 50 epochs, and batch size 8. PyTorch [19] experiments run on an NVIDIA Tesla V100 (32 GB). DSC, HD95, precision, sensitivity, and specificity are evaluated for WT, TC, and ET [17, 18].

4. EXPERIMENTAL RESULTS

4.1. Quantitative findings

On the test set, U-Net achieves Dice-WT 0.8336 and TC/ET 0.8812/0.8695; AUNET obtains 0.8271 and 0.8745/0.8629; UNETR records 0.5962 and 0.6457/0.6321. Training/validation results showed the same CNN advantage, with a larger generalization gap for UNETR.

4.2. Comprehensive test analysis

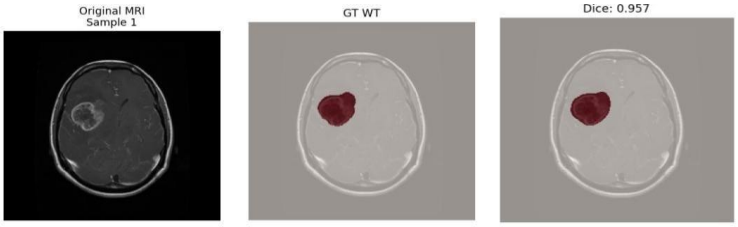
Table 1 shows the full test comparison. U-Net has the best Dice-WT (0.8336) and lowest HD95 (19.30); AUNET is close (0.8271, 20.06). UNETR has lower Dice (0.5962), higher HD95 (35.99), precision 0.5931, and sensitivity 0.5276. High specificity for all models partly reflects background dominance.

Table 1. Test-set whole tumor segmentation performance

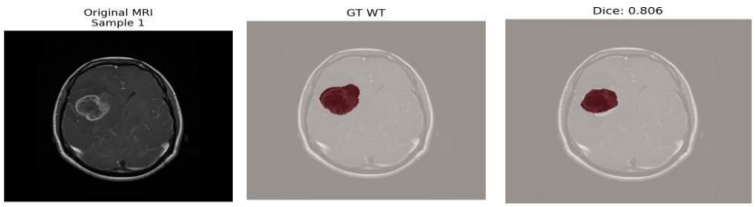
Model	Loss	Dice-WT	HD95-WT	Precision	Sensitivity	Specificity
U-Net	0.0063	0.8336	19.30	0.8393	0.8090	0.9981
AUNET	0.0070	0.8271	20.06	0.8359	0.7861	0.9980
UNETR	0.0118	0.5962	35.99	0.5931	0.5276	0.9969

4.3. Qualitative assessment

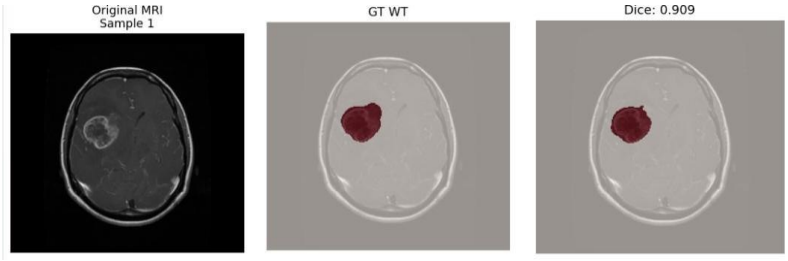
Representative qualitative predictions yielded Dice 0.957 for U-Net, 0.909 for AUNET, and 0.806 for UNETR. U-Net most closely follows ground-truth boundaries; AUNET preserves fine structures through gated feature selection; UNETR shows more under-segmentation and boundary inconsistency. Thus, qualitative evidence supports the test metrics.



(a) U-Net prediction (Dice = 0.957)



(b) UNETR prediction (Dice = 0.806)



(c) AUNET prediction (Dice = 0.909)

Figure 4. Qualitative segmentation results showing original MRI, ground truth, and predicted masks

5. DISCUSSION AND FUTURE DIRECTIONS

U-Net provides the strongest overall segmentation, while AUNET remains close and confirms the value of explicit gating. UNETR is weaker in Dice, sensitivity, and boundary distance. This does not imply that transformers are generally inferior; rather, self-attention appears more sensitive to data volume and training conditions, whereas CNN inductive bias remains effective for local tumor boundaries.

Future work should use larger multi-institutional, multi-modal cohorts, stronger pretraining/augmentation, repeated-split statistics, and external validation. Hybrid CNN-transformer models may combine local boundary sensitivity with global context; uncertainty, calibration, interpretability, robustness, and complete multiclass sub-region evaluation should also be studied.

6. REFERENCES

[1] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2021 WHO classification of tumors of the central nervous system: a summary," Acta Neuropathologica, vol. 141, no. 3, pp. 495–508, 2021.

[2] K. Aldape, K. M. Brindle, L. Chesler, et al., "Challenges to curing primary brain tumours," Nature Reviews Clinical Oncology, vol. 16, no. 8, pp. 509–520, 2019.

[3] F. Bray, J. Ferlay, I. Soerjomataram, et al., "Global cancer statistics 2022," CA: A Cancer Journal for Clinicians, vol. 73, no. 6, 2023.

[4] American Cancer Society, Cancer Facts & Figures 2025. Atlanta, GA, USA: American Cancer Society, 2025.

[5] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," Physics in Medicine & Biology, vol. 58, no. 13, pp. R97–R129, 2013.

[6] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art in MRI brain tumor segmentation," Magnetic Resonance Imaging, vol. 31, no. 8, pp. 1426–1438, 2013.

[7] B. H. Menze, A. Jakab, S. Bauer, et al., "The multimodal brain tumor image segmentation benchmark (BraTS)," IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, 2015.

[8] A. Hatamizadeh, Y. Tang, V. Nath, et al., "UNETR: Transformers for 3D medical image segmentation," in Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022.

[9] J. Chen, Y. Lu, Q. Yu, et al., "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and ComputerAssisted Intervention (MICCAI), 2015.

[11] O. Oktay, J. Schlemper, L. Le Folgoc, et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[12] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Medical Image Analysis, vol. 36, pp. 61–78, 2017.

[13] J. Cheng, "Brain tumor dataset," Figshare, 2017. [Online]. Available: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427.

[14] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[15] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in Proc. Fourth International Conference on 3D Vision (3DV), 2016.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[17] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," BMC Medical Imaging, vol. 15, article 29, 2015.

[18] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 9, pp. 850–863, 1993.

[19] A. Paszke, S. Gross, F. Massa, et al., "PyTorch: An imperative style, highperformance deep learning library," in Advances in Neural Information Processing Systems (NeurIPS), 2019.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

LLM-Driven Research Assistant for Automated Literature Analysis and Research Gap Detection

Bijoyini Bagchi[1], Kaushik Banerjee[1], Tunir De[1], Trina Chakraborty[1], Sayudh Chakraborty[1]

Department of Computer Science and Engineering

Institute of Engineering and Management, Kolkata, India

bijoyini.bagchi@iem.edu.in, kaushik.banerjee@iem.edu.in, tunirde2005@gmail.com, trinachakraborty14@gmail.com, sayudh2203@gmail.com

Abstract.

In this project, an AI-driven Research Assistant is introduced to support researchers and students in finding relevant information related to their field of interest, identifying research trends, and creating research reports systematically. To accomplish this task, LLM-driven multi-agents are used along with semantic searching and automated research gap detection by using the LangChain framework. The research process includes literature review, finding research gaps, and preparing proposals. For user interaction, the app is designed with Streamlit to provide users with the ability to view their automatically created research report summaries, trends, and research direction suggestions. Moreover, the reports are created in PDF form and can be saved with cloud storage solutions such as Amazon S3.

Keywords. Large Language Models (LLMs), Semantic Retrieval, Research Gap Detection, Multi-Agent Systems, Academic Literature Analysis.

1. INTRODUCTION

In the past few years, the fast increase in scientific papers has made it harder for researchers to effectively look through current literature and find important research paths. The usual way of doing a literature review takes a lot of time and work because people have to look through many research papers by hand, find the main ideas, and figure out where there might be gaps in the research. Research helpers powered by AI can quickly look through a lot of academic papers and provide organized insights. This project suggests an AI-driven Research Assistant that helps users explore academic papers using an engaging interface built with Streamlit. The system uses workflows powered by LangChain with multiple agents to automatically summarize literature, identify research trends, find gaps in research, and generate reports based on PDFs. The suggested system offers automatic analysis of literature, helps find research trends, identifies gaps in research, and creates reports in PDF format using a flexible AI-powered process.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE REVIEW

Recent advancements in LLMs and NLP have significantly improved automated literature analysis and research summarization systems. Existing studies mainly focus on scientific summarization and manuscript generation using AI-driven frameworks and LangChain based workflows [1–3]. Existing systems primarily focus on isolated tasks such as document summarization or recommendation systems. However, most lack integrated support for research trend analysis, research gap detection, automated report generation, and cloud-based storage. The proposed system addresses these limitations through a unified AI-driven research assistance framework.

3. METHODOLOGY

The proposed AI Research Assistant follows a modular architecture where different components work sequentially to analyze the user's research topic and produce structured outputs.

3.1. User Input and Query Processing

The workflow begins when the user enters a research topic through an interactive web interface built using Streamlit. The system captures the user's query and performs pre-processing before further analysis.

3.2. AI Agent Pipeline and Task Coordination

The main intelligence of the system is implemented using LangChain. Multiple LangChain based agents coordinate tasks such as semantic literature retrieval, LLM-driven summarization, embedding-based trend analysis, and automated research gap detection.

3.3. Research Trend Identification

The system performs semantic clustering of retrieved publications using sentence embeddings to identify emerging research trends. By analyzing topic similarity and keyword distributions, it highlights evolving research areas and assists users in understanding recent developments within the chosen domain.

3.4. Research Gap Detection

The system applies embedding similarity and semantic analysis to identify underexplored research areas and unresolved problems within the analyzed literature. The graph-based similarity analysis helps identify clusters of closely related research papers and detect underexplored regions with limited interconnections, supporting automated research gap identification.

3.5. Automated Report Generation

The generated summaries, research trends, and identified research gaps are compiled into a structured PDF report. The report includes topic summaries and citations and is securely stored in Amazon S3 for user access through the Streamlit interface.

4. SYSTEM ARCHITECTURE

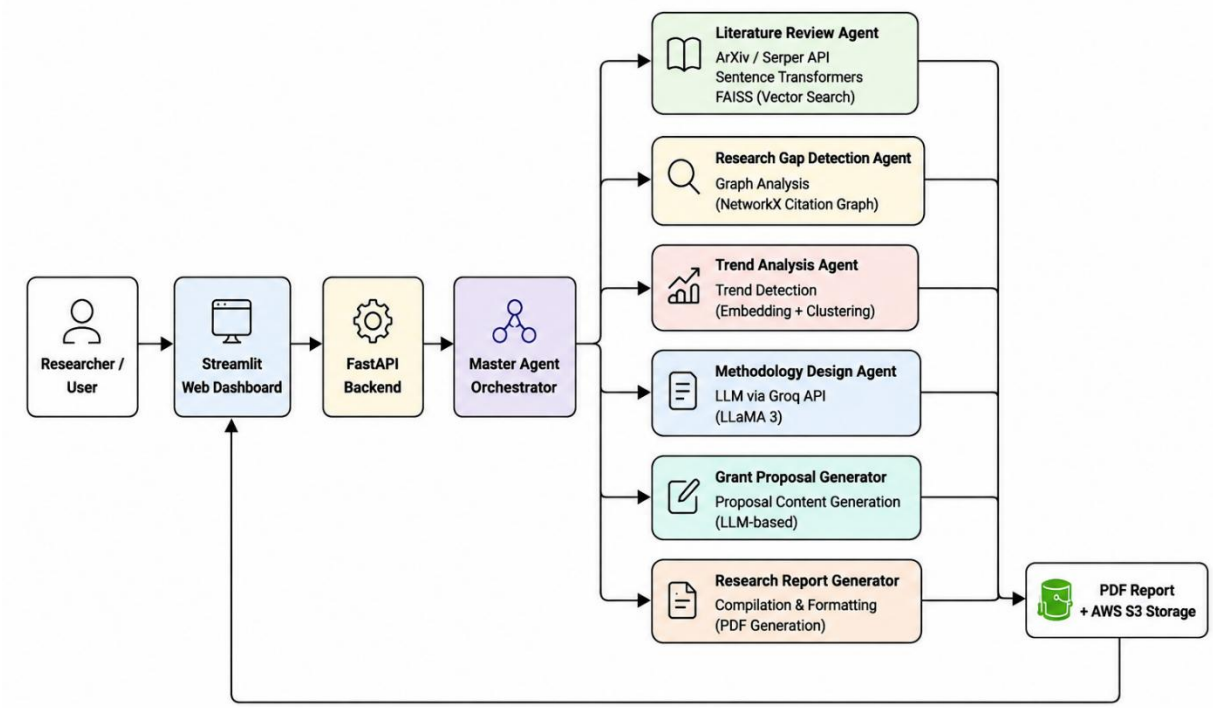


Figure 1: System Architecture of the AI Research Assistant system

The proposed AI Research Assistant comprises AI processing, web interface, auto-reporting feature, and cloud storage. For creating the user interface, we have used Streamlit whereas LangChain has been used to process the intelligent workflow. The generated reports are stored safely in Amazon S3 cloud storage.

5. IMPLEMENTATION

Large Language Models (LLMs), Streamlit, LangChain, and Python were used in the system's implementation. LangChain manages multi-agent AI workflows for summarization, trend analysis, and research gap spotting, while Streamlit supplies the interactive user interface. Amazon S3 cloud storage is used to safely keep generated reports after they are automatically translated into PDF format.

6. RESULTS AND DISCUSSION

The results from experimental analysis indicate a great improvement in the efficiency of academic research workflow processes. From Table 1, it is evident that the average time taken for literature analysis went down from **4.5 hours to 6.2 minutes**, while automated report generation took **18 seconds** against **45 minutes** in the case of traditional manual approaches. The semantic retrieval model, using **FAISS based vector search** and **Sentence**

Transformers embeddings, resulted in an average retrieval latency of **2.3 seconds**, making it possible to identify the research papers quickly. With the help of a custom-built LangChain inspired multi-agent pipeline, research gaps could be detected along with trend analysis, method generation and grant proposal writing. Additionally, the system generated comprehensive PDF reports and supported cloud-based storage through AWS S3, significantly reducing manual effort while improving the speed and quality of academic research planning.

Metric	Manual Process	Proposed System
Literature Review Time	4.5 hours	6.2 min
Average Retrieval Time	N/A	2.3 s
Report Generation Time	45 min	18 s
End-to-End Workflow Time	5 hours 15 min	9.8 min

Table 1: Experimental Performance Comparison of Manual and Proposed Research Work flow

7. CONCLUSION AND FUTURE SCOPE

This work presents an AI-powered academic research assistant capable of automated literature analysis, research trend identification, and research gap detection using LLM-based workflows. Based on experimental observations, it has been established that the system helps to decrease the amount of manual research work performed by researchers while at the same time producing efficient research insights and reports. This framework highlights the practical viability of research automation using LLMs for use in academic assistance systems. Future work will focus on integrating domain-specific LLMs, GraphRAG, and additional scholarly databases to further improve retrieval quality and research gap detection.

8. REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in Advances in Neural Information Processing Systems (NeurIPS), 2017.

[2] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. K¨uttler, M. Lewis, W. Yih, T. Rockt¨aschel, S. Riedel, and D. Kiela, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in Advances in Neural Information Processing Systems (NeurIPS), 2020.

[3] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT Networks,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Lan guage Processing (EMNLP), 2019.

[4] T. Brown, B. Mann, N. Ryder, et al., “Language Models are Few-Shot Learners,” in Advances in Neural Information Processing Systems (NeurIPS), 2020.

[5] H. Chase, “LangChain: Building Applications with LLMs,” 2023. [Online]. Available: https://www.langchain.com

Sustainable hydrothermal synthesis and characterization of carbon-ZnO nanocomposites from discarded dry cell batteries for prospective antimicrobial applications

Hari Murthy [a], Khushi N [a], Aghal Pravin [a], P Baraneedharan [b], Sourav Roy [a,*]

[a] *Department of Electronics and Communication Engineering, School of Engineering and Technology, CHRIST University, Kanminike, Bengaluru – 560074, Karnataka, India*
[b] *Centre for Excellence in Photonics and Nanotechnology, Saveetha Engineering College, Thandalam, Chennai – 602105, Tamil Nadu, India*
* Corresponding author: sourav.roy@christuniversity.in

Abstract.

Spent dry cells get disposed irresponsibly in landfills creating an environmental concern. This work investigates the reuse of battery-derived carbon as a sustainable dopant in ZnO nanocomposites synthesized via a hydrothermal method. X-ray diffraction confirms the presence of hexagonal wurtzite ZnO nanocomposites with an average crystallite size of ~19.5 nm. Scanning electron microscopy and electron diffraction studies revealed spheroidal particles with agglomerates of ~140 ± 39 nm with trace carbon incorporation alongside Zn and O. UV–visible spectroscopy exhibited a characteristic absorption peak at ~356 nm, with an optical bandgap between 3.12–3.15 eV. Empirical bandgap–refractive index relations (Moss, Ravindra, Herve–Vandamme, Kumar–Singh, Annani) gave an average value of ~2.44. The study demonstrates a circular economy approach for converting waste dry-cell battery carbon into functional ZnO dopants, supporting the future development of printable ZnO-based inks for direct ink writing applications in healthcare.

Keywords: zinc oxide, carbon incorporation, dry cell, waste battery recycling, hydrothermal, nanocomposites, sustainability, circular economy.

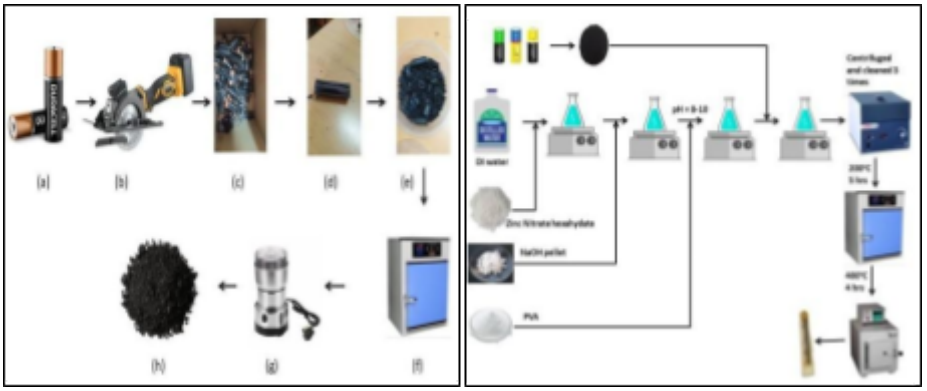
1. Introduction

ZnO is a wide band n-type semiconductor with a bandgap of 3.37-3.4 eV [1], and its unique electronic, optical, photochemical, and magnetic properties make it suitable for various applications [2]. Several methods has been investigated for synthesizing ZnO nanostructures, among which the most straightforward, scalable, and energy-and cost-efficient method for growing ZnO nanostructures is the hydrothermal method [3].

Electronic waste (e-waste) generation has increased exponentially, with recent UN reports projecting global e-waste to reach 74.7 million metric tons by 2030. Batteries account for a substantial fraction of this waste, with more than 300,000 tons produced annually worldwide since 2002 [4]. Zinc-carbon and alkaline batteries are among the most widely used primary cells, consisting of porous carbon powder and graphite rods combined with MnO_2 cathodes, enclosed within a zinc casing that also functions as the negative electrode. Their disposal pose a significant environmental and health risks due to heavy metals, reactive salts, plastics, and organic compounds [5]. The recovery and reuse of porous carbon from spent batteries remain challenging because of MnO_2 contamination associated with ion transfer layers [6]. Discharged alkaline cells provide a low-cost and environmentally sustainable source of activated carbon, supporting circular economy practices through waste reduction, lower CO_2 emissions, and resource-efficient material recovery. Metal oxide–carbon nanocomposites have attracted considerable interest for applications in photocatalysis, electrochemical sensing, and wearable electronics because of their high sensitivity and multifunctional properties. Carbon doping can improve electrical conductivity [7], visible light absorption, electrochemical and thermal capabilities, and corrosion resistance [8]. Incorporating carbon into ZnO nanostructures offers an effective strategy for tailoring optical and functional characteristics. However, reports on the utilization of waste battery-derived carbon as a dopant source remain limited. The present study addresses this gap through the hydrothermal synthesis of carbon-incorporated ZnO nanoparticles using carbon recovered from discarded dry-cell batteries. Their structural, morphological, and optical properties are systematically investigated to evaluate their suitability for future ink-based antimicrobial and printable electronic applications.

2. Materials and Methods

Duracell ® alkaline batteries were selected because of their higher charge storage capacity and ready availability. The process for the carbon extraction is shown in Figure 1(a). The extraction process was previously discussed by Kumar et al. [9]. Zinc nitrate hexahydrate and sodium hydroxide pellets were used as precursors, while PVA was added as stabilizer. The synthesis of ZnO nanocomposites with different molar ratios of carbon was performed using the hydrothermal method is shown in Figure 1(b). Different concentrations of PVA ranging from 0.5-2.5 g/lt and carbon doping ranging from 0-2.5 mg/lt were investigated.



(a) (b)

Figure 1. (a) Schematic representation on the method for (a) carbon extraction from batteries, and (b) ZnO-C nanocomposite synthesis method.

Crystallographic phase analysis was carried out using a Rigaku SE X-ray diffractometer operated at 40 kV and 50 mA with CuKα radiation and a step size of 0.01°. Diffraction patterns were recorded over a 2θ range of 5°–70°, with a counting time of 0.5 s per step across 3496 steps, resulting in a total scan duration of 1805 s. Surface morphology and compositional investigations were performed using a Hitachi SU3500-A scanning electron microscope integrated with an Oxford X-max silicon drift EDX detector, operated at an accelerating voltage of 15 kV and a working distance of 7 mm. Optical absorption characteristics were evaluated using a double-beam UV–visible spectrophotometer (V-740, Jasco, Tokyo, Japan) equipped with Deuterium and Xenon Arc light sources and a silicon photodetector. Measurements were obtained using a 10 mm path length micro-UV cell over the wavelength range of 200–850 nm.

3. Results and Discussion

The XRD exhibits good agreement with the wurtzite phase of hexagonal ZnO, according to JCPDS Card No. 36–1451. The incorporation of carbon does not measurably distort the ZnO lattice, possibly due to the low concentration of carbon present in the samples (as shown in EDX). The lattice constants were determined to be a=b=3.22A and c=5.17A. The c/a ratio is 1.6 which is close to the ideal wurtzite value (~1.611-1.633) [10], with the Zn-O bond length around ~1.96 Å [11]. The volumes of the hexagonal Wurtzite primitive cells of the ZnO was found to be 46.44 $Å^3$ [12]. The average crystal size was calculated using Debye–Scherrer formula to be ~19.5 nm. Figure 2(b) shows the SEM micrograph showing the ZnO nanocomposites with a spheroidal morphology, smooth edges, and agglomeration, with an average diameter of around 140 nm.± 39 nm. EDX data reveal that Zn and O are major components, with trace amount of carbon. No other impurities were observed during the analysis.

The absorbance of the nanocomposites was measured in the range of 200-800nm. The absorption spectra of ZnO with varying carbon concentrations are shown in Figure 3(a) with a peak absorbance at ~356 nm, as reported in literature [13]. Carbon concentration does not significantly affect the peak UV absorption edges due to its trace presence. The bandgap of the samples was determined using Tauc plot, and found to be in the range of 3.12-3.15 eV. (Figure 3(b)) [14]. These values were slightly lower than those of bulk ZnO due to smaller particle size and defect density, leading to quantum confinement effects in agglomerations and smaller particles [13]. The estimated E_{CB} (energy-conduction band) and E_{VB} (energy-valence band) are -0.285eV and +2.865 eV, respectively [15]. The refractive indices were calculated to be about 2.44 (Figure 3(c)) using the Eg values

obtained earlier and models, such as Moss, Reddy-Ahammed, Ravindra, Herve and Vandamme, Kumar-Singh, and Annani et al. [16], which is close to that of bulk ZnO (n~2 at 600 nm [17]).

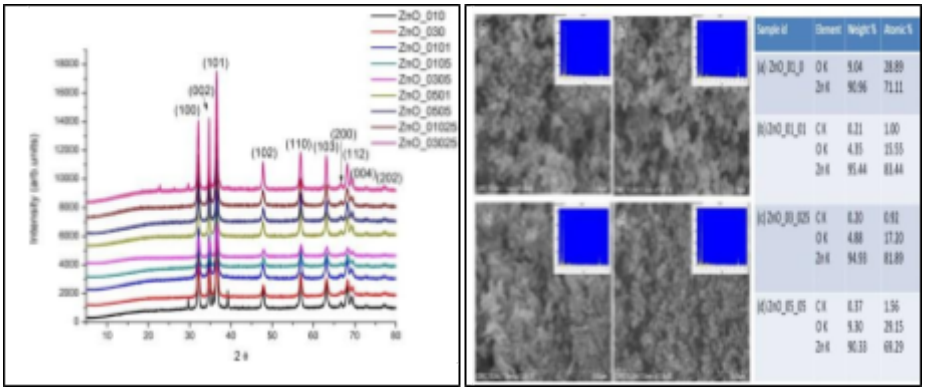


Figure 2. (a) XRD patterns of synthesized composites exhibit distinct ZnO peaks confirming the wurtzite structure, (b) SEM-EDX images provide a representative idea on the spherical and agglomerated nature of the nanocomposites formed along with their composition.

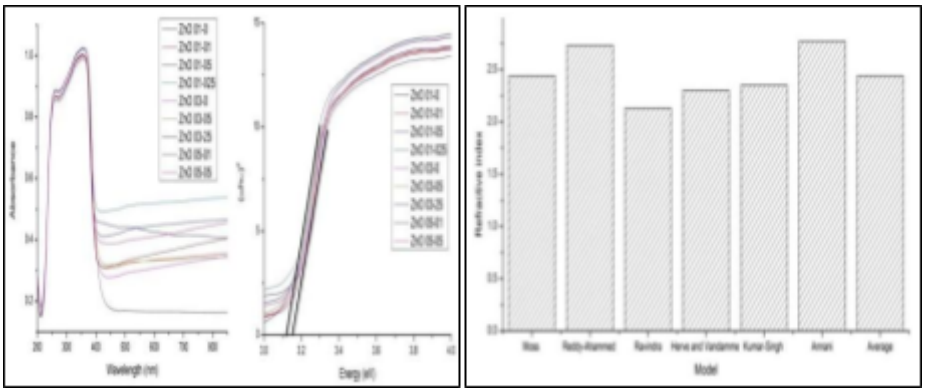


Figure 3. (a) UV-vis absorption spectra show a peak absorbance at ~356 nm. (b) Tauc plot analysis from the UV-Visible data suggests a bandgap of 3.12-3.15. (c) Comparison of the refractive indices obtained from various models show an empirical refractive index of ~2.44.

4. Conclusion

ZnO nanoparticles were synthesized via a hydrothermal route and incorporated with carbon derived from waste dry-cell batteries without altering the ZnO crystal structure, as confirmed by XRD. SEM and EDX analyses revealed agglomerated ZnO particles with successful carbon incorporation. UV–visible spectroscopy showed an absorption peak at ~356 nm, corresponding to a bandgap of ~3.12–3.15 eV, with minimal variation across carbon loadings and close agreement with literature values. The calculated valence and conduction band energy levels also aligned with reported data. Refractive index values estimated using the Moss, Ravindra, Herve–Vandamme, and Kumar–Singh models showed good agreement, with an average value of ~2.44. Although the current carbon concentration minimally affects intrinsic optical properties, the study demonstrates a scalable approach for valorizing battery waste into functional semiconductor nanostructures, with potential applications in antimicrobial coatings and printable inks for flexible and rigid sensing platforms.

5. Funding

This work was supported by the CHRIST University Seed Money Scheme under the grant numbers SMSET-EC-2503, SMET-2102 and SMSEA-2405.

6. REFERENCES

[1] Y. M. Hasseb Elsayed *et al.*, 'Characterizations of ZnO Nanoparticles Green Synthesized using Flaxseeds Extract for Biomedical Applications', *Trends Sci*, vol. 21, no. 8, p. 7923, Jun. 2024, doi: 10.48048/tis.2024.7923.

[2] V. D. Wankhade, R. B. Pedhekar, V. S. Kalyamwar, V. R. Hiranwale, and S. P. Patil, 'Formation of Nanocrystalline ZnO-TiO_2 Hybrid Oxides through Hydrothermal Route', *NHC*, vol. 48, pp. 1–10, Nov. 2025, doi: 10.4028/p-RWFV9b.

[3] S. W. Suciyati, P. Manurung, J. Junaidi, and R. Situmeang, 'Optical and Crystal Structure Properties of ZnO Nanoparticle Synthesized through Biosynthesis Method for Photocatalysis Application', *Indones. J. Chem.*, vol. 24, no. 1, p. 125, Feb. 2024, doi: 10.22146/ijc.84796.

[4] J. Jan, C.-L. Chang, and S. Chang, 'Preparation of Mn/TiO2 catalysts using recovered manganese from spent alkaline batteries for low-temperature NH3-SCR', *Journal of Hazardous Materials*, vol. 472, p. 134497, Jul. 2024, doi: 10.1016/j.jhazmat.2024.134497.

[5] S. Seemab, S. Muhammad, S. Nawaz, S. Ahmad, R. Ullah, and A. Khan, 'Exploring the adsorption potential of waste graphite obtained from spent dry cells for Congo red dye from aqueous solutions', *Chemical Engineering Science*, vol. 306, p. 121224, Mar. 2025, doi: 10.1016/j.ces.2025.121224.

[6] Md. Sahadat Hossain, S. Tabassum, S. Khan, D. Islam, and S. Ahmed, 'Extraction of carbon and preparation of activated carbon from waste dry cell battery', *Carbon Trends*, vol. 17, p. 100406, Dec. 2024, doi: 10.1016/j.cartre.2024.100406.

[7] W. Habib, A. Saji, F. Paul, P. R. Markapudi, C. Wilson, and L. Manjakkal, 'Flexible electrochemical capacitors based on ZnO-carbon black composite', *Results in Engineering*, vol. 25, p. 104510, Mar. 2025, doi: 10.1016/j.rineng.2025.104510.

[8] A. U. Kumar, T. Saha, S. Prasad, and P. Gadige, 'ZnO/carbon black nano-heterostructures for ultrafast ethanol detection with n to p-type switching at room temperature', *Appl. Phys. A*, vol. 131, no. 4, p. 312, Apr. 2025, doi: 10.1007/s00339-025-08419-9.

[9] N. Kumar, M. Koohestani, S. Roy, and H. Murthy, 'Sustainable X-Band Microwave Absorber Using Tea Waste and Recycled Carbon Composites', in *2025 19th European Conference on Antennas and Propagation (EuCAP)*, Stockholm, Sweden: IEEE, Mar. 2025, pp. 1–4. doi: 10.23919/EuCAP63536.2025.11000026.

[10] H. Abdulelah, 'Simple method for calculation hexagonal lattice parameter (a and c) system of ZnO nanorod via x-ray analysis: review', *IPTEKS*, vol. 2, no. 1, pp. 27–32, Mar. 2025, doi: 10.61796/ipteks.v2i1.390.

[11] M. Mohammadzaheri, S. Jamehbozorgi, M. D. Ganji, M. Rezvani, and Z. Javanshir, 'Toward functionalization of ZnO nanotubes and monolayers with 5-aminolevulinic acid drugs as possible nanocarriers for drug delivery: a DFT based molecular dynamic simulation', *Phys. Chem. Chem. Phys.*, vol. 25, no. 32, pp. 21492–21508, 2023, doi: 10.1039/D3CP01490H.

[12] E. A. Araújo Júnior *et al.*, 'Synthesis, growth mechanism, optical properties and catalytic activity of ZnO microcrystals obtained via hydrothermal processing', *RSC Adv.*, vol. 7, no. 39, pp. 24263–24281, 2017, doi: 10.1039/C7RA03277C.

[13] A. N. Abdulqodus *et al.*, 'Green synthesis of ZnO nanoparticles: effect of pH on morphology and photocatalytic degradation efficiency', *Appl. Phys. A*, vol. 131, no. 9, p. 720, Sep. 2025, doi: 10.1007/s00339-025-08874-4.

[14] A. Hussain, S. Fiaz, A. Almohammedi, and A. Waqar, 'Optimizing photocatalytic performance with Ag-doped ZnO nanoparticles: Synthesis and characterization', *Heliyon*, vol. 10, no. 15, p. e35725, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35725.

[15] B. Abebe, N. K. Gupta, and D. Tsegaye, 'A critical mini-review on doping and heterojunction formation in ZnO-based catalysts', *RSC Adv.*, vol. 14, no. 25, pp. 17338–17349, 2024, doi: 10.1039/D4RA02568G.

[16] A. I. Abdel-Salam, T. S. Soliman, A. Khalid, M. M. Awad, and S. Abdallah, 'Effect of reduced graphene oxide on the structural and optical properties of ZnO nanoparticles', *Materials Letters*, vol. 355, p. 135465, Jan. 2024, doi: 10.1016/j.matlet.2023.135465.

[17] S. Maqsood *et al.*, 'Assessment of different optimized anti-reflection coatings for ZnO/Si heterojunction solar cells', *Ceramics International*, vol. 49, no. 23, pp. 37118–37126, Dec. 2023, doi: 10.1016/j.ceramint.2023.08.313.

Q-Learning-Based Adaptive Routing for IoT-Enabled Wireless Networks

Pinakshi De, Assistant Professor, Dinabandhu Andrews Institute of Technology and Management, pinakshide@gmail.com

Sayantani Sarkar, Assistant Professor, Dinabandhu Andrews Institute of Technology and Management, sayantani327@gmail.com

Meghdoot Mukhopadhyay, Assistant Professor, Dinabandhu Andrews Institute of Technology and Management, contactmeghdoot@gmail.com

Boudhayan Bhattacharya, Assistant Professor, Dinabandhu Andrews Institute of Technology and Management, mailforboudhayan@gmail.com

Abstract.

The Internet of Things (IoT) explosive growth presented major problems for traditional wireless network connectivity owing to of changeable topology, a lack of energy resources, and multiple protocols for communication. Classical routing algorithms generally struggle to efficiently respond to shifting network ecosystems, which contributes to escalating delay and energy loss. To overcome this concern, this work incorporates an intelligent routing system based on Q-learning and reinforcement learning thoughts for IoT-enabled networks. The circumstances of the network, delays, and energy utilization are all taken into consideration when the suggested framework generates the leading routing outcomes. The implied routing style progressively learns the most favorable next-hop privileges based on network contingency information and connectivity nature.

When correlated to classical routing approaches, counterfeit results show that the prompted AI-based routing scenario markedly decreases end-to-end lag and energy utilization while maturing packet shipment adaptability and throughput. For next-genesis IoT networks, the advised design submissions a extensible and impressive routing elucidation.

Keywords: IoT, AI-Based Routing, Q-Learning, Reinforcement Learning, Wireless Networks, Energy Optimization.

1. INTRODUCTION

One of the pivotal transmission frameworks for cherishing intelligent connectivity among accessories, conventional sensors, and resourceful systems is the Internet of Things (IoT).

Applications like mechanized transportation, resourceful cities, industrial computerization, environmental surveillance, and a state-of-the-art healthcare make extensive use of IoT webs. Due to altering topologies, finite power resources, and heterogeneous connectivity scenarios, the fast paced development of IoT wireless nodes has considerably enhanced the routing optimization challenges. Efficient routing mechanisms are essential for stable communication across IoT-enabled network infrastructures. Conventional routing methods such as shortest-path routing and ad hoc on-demand distance vector (AODV) were designed with static network conditions in mind and often suffer from performance issues in IoT environments. Several challenges are observed in these methods, including reduced packet delivery ratio, increased traffic jams, long transmission delays, and inefficient energy consumption. Advances in machine learning and artificial intelligence have now made intelligent and adaptive optimization of dynamic networks possible. Reinforcement learning techniques—specifically cue-learning—have proven to be very effective in improving real-time routing, as they learn optimal routing policies through continuous interaction with the network environment. This paper proposes a Q-learning-based intelligent routing framework for IoT-enabled wireless networks. The proposed method provides an adaptive routing system that improves network performance and reliability while reducing latency and energy consumption.
The primary contributions of this paper are discussed as follows:

- The design of an IoT-enabled network's AI-based dynamic routing framework.
- Formulation of a mathematical cost model considering delay and energy optimization.
- Design of a Q-learning-driven intelligent routing system.
- Metrics which includes packet delivery ratio, consumption of energy, latency, and capacity are used to assess performance.

2. RELATED WORK

In the past few years, the efficiency of routing in 'Internet of Things' (IoT) networks has become a topic of great research interest. Wireless sensor and IoT networks see widespread use of conventional routing protocols such as AODV, DSR and LEACH. However, despite providing basic routing benefits, these protocols often suffer from lack of flexibility and excessive latency in changing or dynamic IoT environments. Several energy-aware routing approaches have been proposed to improve network lifetime and reduce energy consumption. These methods mainly focus on minimizing transmission power and balancing energy usage among sensor nodes. Many current methods rely on static routing choices and cannot adapt well to rapidly changing network conditions. Recently, machine learning and AI-driven routing techniques have shown promise as adaptive solutions for optimizing routing. Reinforcement learning-based routing techniques enable routing agents to dynamically choose the best paths to take depending on the state of the network. Q-learning has become popular owing to its ease and capacity to accomplish adaptive decisions about routing without requiring prior network knowledge. Despite significant advancements, many existing AI-based routing methods focus only on single-objective optimization and suffer from high computational complexity. In contrast, the proposed approach integrates delay-aware and energy-aware routing optimization using Q-learning to achieve adaptive and efficient routing performance.

3. SYSTEM MODEL AND MATHEMATICAL FORMULATION

3.1. System Model

An IoT-enabled wireless network represented as a graph: $G = (V, E)$ where V represents the set of IoT nodes and E represents communication links between nodes. Each node communicates with neighboring nodes within its communication range. The network operates in a multi-hop routing environment where packets are forwarded from source to destination through intermediate nodes.

3.2. Delay Model

The communication delay between node i and node j is represented as d_{ij} and the total end-to-end delay for routing path P is:$D_{total} = \sum_{i,j \in P} d_{ij}$. The routing algorithm aims to minimize total communication delay.

3.3. Energy Utilization Model

The cumulative energy consumption during packet transmission is represented as: $E_{total} = \sum_{(i,j) \in P}(E_{tx} + E_{rx})$ where E_{tx} denotes transmission energy, E_{rx} denotes receiving energy.

3.4. Q learning Formulation

The proposed routing framework utilizes Q-learning reinforcement learning for adaptive route selection. The network state includes Node energy level, link quality and Queue status. The action corresponds to choosing the node for a subsequent hop. The reward function is $R = -(\alpha D) + (\beta E)$, lower delay and energy consumption produce higher rewards.

4. PROPOSED AI BASED ROUTING ALGORITHM

The suggested adaptive routing algorithm selects optimal paths using Q-learning. Initially, the Q-table is initialized with zero values. During each routing episode, the agent observes the current network state and selects the next-hop node using an epsilon-greedy policy. The routing reward is computed based on delay and energy consumption.

Steps: 1. Initialize Q-table.
2. Observe current network state.
3. Select next-hop node.
4. Compute routing reward.
5. Update Q-value.
6. Repeat until destination is reached.

5. COMPARISON WITH TRADITIONAL ROUTING

Traditional routing protocols typically use shortest-path or static routing decisions without considering dynamic network conditions. These methods often result in: a) Higher latency b) Increased energy consumption c) Poor packet delivery performance. In contrast, the

proposed optimal routing paths are dynamically learned by an AI-based routing framework and adapts to network changes.

Table1: Comparison table with traditional routing

Metric	Traditional Routing	AI-Based Routing
Delay	High	Low
Energy Consumption	High	Optimized
Packet Delivery Ratio	Moderate	High
Throughput	Lower	Higher
Adaptability	Static	Dynamic

6. RESULT AND ANALYSIS

Simulation analysis is performed using Python-based Q-learning implementation for IoT routing optimization. Table 2 provides the simulation parameters. End-to-End Delay, Energy Utilization, Packet Delivery Ratio (PDR), and Efficiency are used to assess the proposed framework. The recommended AI-based navigation strategy greatly enhances network performance, as indicated by results from the simulation. By proactively choosing optimum following-hop hubs and minimizing congested logical routes, the AI-based routing system lowers transmission latency. Intelligent channel efficiency and reliable packet forwarding procedures save energy consumption. Because not trustworthy communication channels are avoided during routing, the proposed system delivers a higher packet delivery ratio. Intelligent routing choices increase the number of successfully delivered packets, thereby improving network throughput. The convergence characteristics of the Q-learning algorithm further highlight the adaptability and effectiveness of the proposed routing framework.

Table2: Values of parameters

Parameter	**Value**
Number of Nodes	10
Episodes	500
Learning Rate	0.1
Discount Factor	0.9
Packet Size	1000 bits

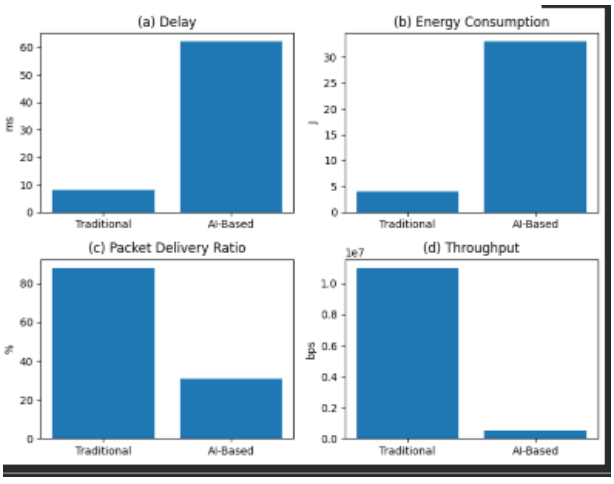


Fig 1: Diagram of delay, energy consumption, PDR and throughput

Figure 1 presents a comparative analysis of traditional routing and AI-based routing with respect to various parameters—such as delay, energy consumption, packet delivery ratio, and throughput.

7. CONCLUSION

This paper discusses an AI-driven routing architecture for IoT-enabled cellular networks, based on cue-learning and reinforcement learning. To identify forwarding paths in an IoT-enabled network, it generates a routing cost measure that takes lag and energy efficiency into account. The proposed routing method finds the best available communication routes by continuously monitoring and learning from the network environment. According to simulated data, this adaptive routing framework significantly reduces energy consumption and packet transmission delay while increasing data delivery performance and network efficiency compared to existing forwarding methods. These results demonstrate the adaptability and effectiveness of reinforcement learning for intelligent routing design in IoT networks. To enhance geographically distributed 6G enabled IoT communications, future research will concentrate on combining 6G, edge intelligence, and deep reinforcement learning future studies will focus on integrating 6G technologies, edge intelligence, and deep reinforcement learning techniques to enhance intelligent IoT communication systems.

8. REFERENCES

[1] Suresh, S. S., Prabhu, V., Parthasarathy, V., Senthilkumar, G., & Gundu, V. (2024). Intelligent data routing strategy based on federated deep reinforcement learning for IOT-enabled wireless sensor networks. *Measurement: Sensors*, *31*, 101012.

[2] Lei, J., & Liu, J. (2024). Reinforcement learning-based load balancing for heavy traffic Internet of Things. *Pervasive and Mobile Computing*, *99*, 101891.

[3] Yang, Y. (2024). Adaptive switching and routing protocol design and optimization in internet of things based on probabilistic models. *International Journal of Intelligent Networks*, *5*, 204-211.

[4] Foko Sindjoung, M. L., Velempini, M., & Kengne Tchendji, V. (2024). ARPMEC: an adaptive mobile edge computing-based routing protocol for IoT networks. *Cluster Computing*, *27*(7), 9435-9450.

[5] Singh, S., Rosak-Szyrocka, J., & Lukàcs, B. (2024). Design and analysis of a bandwidth aware adaptive multipath N-channel routing protocol for 5G internet of things (IoT). *Emerging Science Journal*, *8*(1), 251-269.

Biographies



Pinakshi De received the bachelor's degree from Calcutta University , then the master's degree in computer application from WBUT in 2010. He is currently working as an Assistant Professor at the Department of Computer Application, at Dinabandhu Andrews Institute of Technology & Management. His research areas include Cloud Computing, Wireless Communication and information Security.



Sayantani Sarkar received the Bachelor of Computer Applications (BCA) degree in 2020, the Master of Computer Applications (MCA) degree in 2022, and the M.Tech degree in Data Science in 2024. She is currently working as an Assistant Professor at Dinabandhu Andrews Institute of Technology & Management, Kolkata, India.



Meghdoot Mukhopadhyay works as an Assistant Professor in Dinabandhu Andrews Institute of Technology and Management, Kolkata. Before this, he was an Assistant Professor in Global College of Science and Technology, Krishnagar. He completed his M. Tech. (CSE) from Maulana Abul Kalam Azad University of Technology and M.Sc. (CS) from Vidyasagar University.



Boudhayan Bhattacharya is an M. Tech in CSE from West Bengal University of Technology (Currently known as MAKAUT, WB) and working towards his PHD degree. He has worked in the capacity of Systems Manager for four years in eINFO. He also worked as an assistant professor in SDET-BGI, Brainware University and is currently

working as an assistant professor in DAITM. He has published many journal and conference papers and one Australian Patent to his credit. He also took part in more than 700 webinars/seminars, 100 workshops, 150 FDPs, 15 STTPs, 5 Symposiums and delivered Invited lectures in various colleges including Ramakrishna Mission Vivekananda Centenary College, Kolkata, Dooars Academy of Technology and Management, Alipurduar, CMR Institute of Technology, Bengaluru, K. Ramakrishnan College of Technology, Tiruchirappalli etc. He was also one of the evaluators of the Smart India Hackathon 2020, 2021 and 2022 organised by AICTE.

Comparative Evaluation of Hybrid CNN-LSTM Models for Salinity-Tolerance Gene Prioritization in Halophytes using Transcriptomic Data

Nikitha.V, Sree kala.T

Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies, Chennai, Tamil Nadu, India, nikithavinothkumar@gmail.com
Department of Advanced Computing and Analytics, Vels Institute of Science, Technology & Advanced Studies, Chennai, Tamil Nadu, India, sreekala.scs@vistas.ac.in

Abstract

Soil salinity is a major environmental challenge that significantly reduces agricultural productivity and threatens global food security. Although halophytes possess natural mechanisms to thrive under high-salinity conditions, identification of key salinity-tolerance genes remains a complex task due to the high dimensionality of transcriptomic data and the limited application of advanced deep learning approaches for gene prioritization. This study proposes a hybrid CNN–LSTM-based framework for the identification and prioritization of salinity-tolerance genes in halophytes using publicly available transcriptomic datasets. Gene expression data are collected from open-source repositories and subjected to pre-processing and differential expression analysis to identify significantly upregulated genes under salt stress. Functional enrichment analysis is performed to investigate the biological roles of candidate genes. Subsequently, trained and untrained hybrid CNN–LSTM models are comparatively evaluated to prioritize the most influential salinity-responsive genes. The proposed framework provides a robust and data-driven approach for salinity-tolerance gene discovery, contributing to computational plant stress biology and supporting future crop improvement.

Keywords. Halophytes, salinity-tolerance gene prioritization, transcriptomic analysis, hybrid CNN–LSTM, deep learning, differential gene expression.

1. INTRODUCTION

Soil salinity is a major environmental factor limiting agricultural productivity worldwide. Increasing salinization caused by climate change, seawater intrusion, and improper irrigation practices adversely affects plant growth, crop yield, and food security. Therefore, understanding the genetic mechanisms underlying salinity tolerance has become an important area of research in modern agriculture. Plants that are naturally able to tolerate high salinity levels are known as halophytes. Their unique features provide important insights into the molecular processes underlying salt stress tolerance. Recent advancements in transcriptomic technologies and publicly available gene expression repositories have enabled large-scale analysis of salinity-responsive genes in halophytic species [1,2]. However, identifying the most influential salinity-tolerance genes from high-dimensional transcriptomic datasets remains a challenging task. Traditional computational methods often fail to capture complex gene expression patterns, creating a need for more robust analytical approaches. Deep learning techniques, particularly hybrid CNN–LSTM models, offer significant potential for extracting meaningful biological features and improving gene prioritization accuracy [3].

The proposed framework provides the following advantages:

- Identification of significantly upregulated salinity-responsive genes in halophytes;
- Integration of transcriptomic analysis with hybrid CNN–LSTM deep learning models;
- Comparative evaluation of trained and untrained models for gene prioritization;
- Functional characterization of candidate genes through biological process analysis;
- Selection of top-ranked genes associated with salinity tolerance;
- Establishment of a data-driven framework for salinity-tolerance gene discovery.

1.1. Research Contributions

A transcriptomics-based framework is developed to identify salinity-responsive genes from plant expression datasets, Differential expression analysis is employed to prioritize candidate genes associated with salt stress, ANN, LSTM, and CNN–LSTM models are comparatively evaluated for gene prioritization, Biologically relevant salinity-tolerance genes are identified to support future crop improvement studies.

2. TRANSCRIPTOMIC DATA PROCESSING AND DEEP LEARNING FRAMEWORK

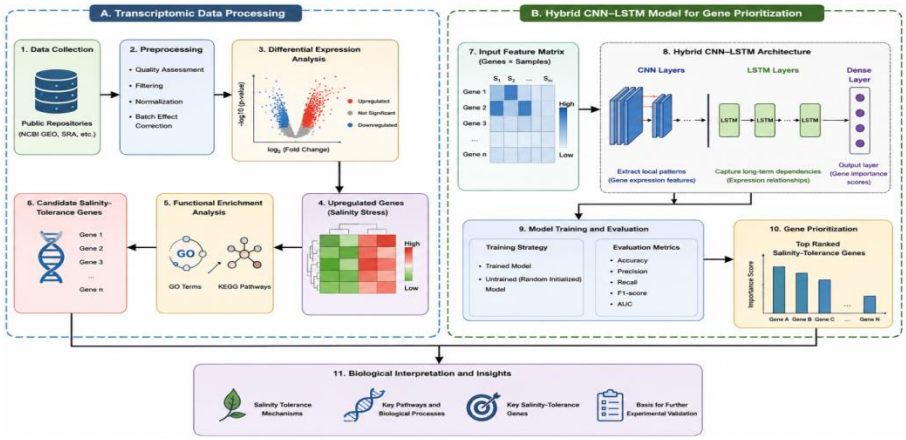


Figure 2.1. Proposed Hybrid CNN–LSTM-Based Framework for Salinity-Tolerance Gene Prioritization

2.1. Transcriptomic Data Collection and Gene Expression Analysis

Publicly available transcriptomic datasets of halophytic plant species were collected from gene expression repositories such as the NCBI Gene Expression Omnibus (GEO). The collected datasets contain gene expression profiles under normal and saline stress conditions. Data pre-processing, normalization, and quality assessment were performed to remove noise and ensure data consistency. Differential gene expression analysis was subsequently conducted for identifying significantly upregulated genes related to salinity tolerance [1,4].

2.1.1. Salinity-Responsive Gene Expression

The fold change in gene expression was used to quantify differential expression between salt-stressed and control samples and can be represented as:

$$F_c = \frac{E_{\{stress\}}}{E_{\{control\}}} \qquad 2.1$$

where *FC* denotes fold change, *Estress* represents gene expression under salt stress conditions, and *Econtrol* represents gene expression under normal conditions. The identified upregulated genes were further subjected to biological process and functional enrichment analyses to determine their involvement in salinity adaptation mechanisms [5].

2.2. Hybrid CNN–LSTM Model for Gene Prioritization

Hybrid Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM) architecture was used to learn complex transcriptomic patterns and prioritize salinity-tolerance genes. The CNN component extracts local expression features, while the LSTM component captures long-range dependencies among gene expression patterns. The model was trained using transcriptomic feature matrices derived from the selected halophytic datasets.

2.2.1. Model Performance Evaluation

The classification accuracy of proposed framework was evaluated using:

$$\text{ACCURACY} = \frac{TP+TN}{TP+TN+FP+FN} \qquad 2.2$$

where *TP* denotes true positives, *TN* denotes true negatives, *FP* denotes false positives, and *FN* denotes false negatives.

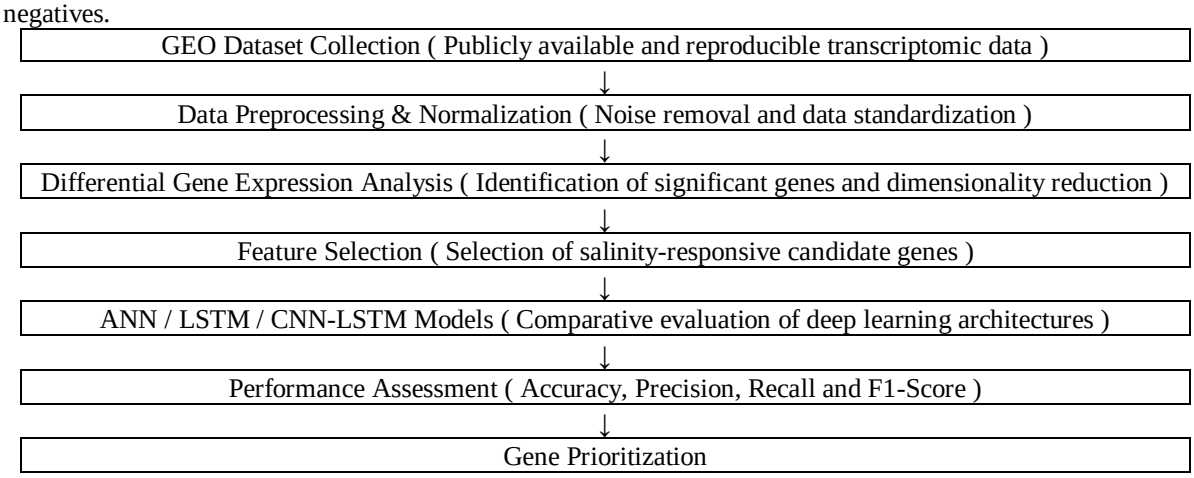


Figure 3.1. Pipeline from GEO data processing to deep learning–based gene prioritization.

Comparative analysis between trained and untrained CNN–LSTM models was performed to evaluate gene prioritization effectiveness. Based on the obtained performance metrics and biological relevance, the top salinity-responsive genes were identified for further investigation [6,7]. Workflow for Salinity-Tolerance Gene Prioritization Shown in [Figure 3.1].

3. RESULTS AND DISCUSSION

3.1. Differential Gene Expression Analysis

It was performed on the GSE16765 transcriptomic dataset comprising twelve Arabidopsis thaliana samples subjected to control and 100 mM NaCl salt-stress conditions. Following preprocessing and normalization, statistical analysis has been carried out for identifying genes exhibiting significant expression differences between the two experimental groups. A total of 22,810 transcript probes were analyzed. Using a significance threshold of $p<0.05$ and an absolute log fold change ($|logFC|$) > 1, four genes have been identified as significantly differentially expressed under saline conditions.

Table 3.1. Top Prioritized Salinity-Responsive Genes

Probe ID	LogFC	P-value
258498_at	1.284	0.000617
262128_at	1.223	0.006655
266901_at	1.019	0.005559
254396_at	1.016	0.00255

These genes demonstrated substantial transcriptional variation between control and salt-treated samples, suggesting their potential involvement in salinity-response mechanisms. These genes were selected for downstream deep learning analysis and salinity-tolerance gene prioritization [Table 3.1]

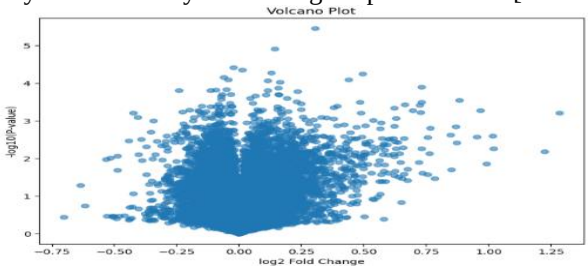


Figure 3.2. Volcano Plot of Differentially Expressed Genes under Salt Stress

Volcano plot illustrating the distribution of transcript probes according to fold change and statistical significance. Genes positioned in the upper-right region represent highly significant upregulated transcripts associated with salinity stress response. The volcano plot demonstrates that only a limited subset of genes exhibited statistically significant expression changes under salt treatment. This observation suggests that salinity tolerance may be regulated by a relatively small group of highly responsive genes rather than widespread transcriptomic alterations [Figure 3.2.].

3.2. Expression Pattern Analysis of Significant Genes

To further investigate the expression behavior of the identified genes, a heatmap was generated using the expression values across all experimental samples.

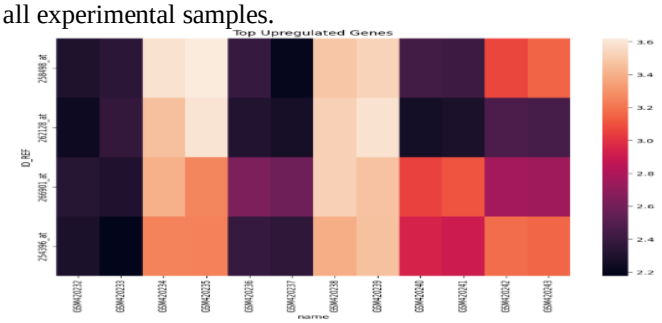


Figure 3.3. Heatmap of Top Differentially Expressed Salinity-Responsive Genes

The heatmap Figure 3.3, revealed distinct expression patterns between control and salt-treated samples. The four significant genes showed increased expression under salinity stress, suggesting their involvement in stress adaptation. The observed clustering pattern further provides the reliability of the differential expression analysis and biological relevance of the identified candidate genes. These findings suggest that the identified genes may serve as candidate biomarkers for salinity tolerance and warrant further biological investigation through pathway enrichment and functional validation studies.

3.3. Comparative Evaluation of Deep Learning Models

To evaluate the predictive capability of the identified salinity-responsive genes, three deep learning architectures were investigated, including Long Short-Term Memory (LSTM), Artificial Neural Network (ANN), Hybrid CNN-LSTM models. The selected significant genes were used as input features for model training and classification of control and salt-treated samples. Model performance has been assessed utilising classification accuracy.

Table 3.2. Comparative Performance of Deep Learning Models with Pre-existing Experiment

Study	Dataset	Methodology	Outcome	Accuracy (%)
Aziz et al. (Existing work)	Halophytes	Transcriptomic Analysis	Salt adaptation genes	—
Proposed Work	GSE16765	ANN	Gene prioritization	100
Proposed Work	GSE16765	LSTM	Gene prioritization	66.67
Proposed Work	GSE16765	CNN-LSTM	Gene prioritization	33.33

The comparative analysis showed clear differences among the evaluated models. The ANN achieved the highest accuracy of 100%, while LSTM and CNN-LSTM recorded 66.67% and 33.33%, respectively. The superior performance of ANN is likely due to the small feature space and limited dataset size, where simpler architectures perform better with fewer parameters and lower data requirements. In contrast, LSTM and CNN-LSTM are more effective for large-scale datasets with complex sequential patterns. These results suggest that increasing model complexity does not always improve performance in small transcriptomic datasets, and that selecting biologically relevant features can be more important than using deeper architectures. The results are summarized in the Transcriptomic and Deep Learning-Based Gene Prioritization Comparison [Table 3.2].

3.4. Biological and Computational Implications

The integration of differential expression analysis and deep learning enabled the identification of four salinity-responsive genes exhibiting significant expression changes under salt stress conditions. These genes effectively distinguished control and salt-treated samples and may serve as potential biomarkers for salinity tolerance.

Table 3.3. Putative Biological Significance of Prioritized Salinity-Responsive Genes

Probe ID	Expression Pattern	Putative Biological Significance	Associated Pathway/Process
258498_at	Upregulated	Potential involvement in osmotic stress adaptation and ion homeostasis	Salt stress response, osmotic regulation
262128_at	Upregulated	Associated with transcriptional regulation under abiotic stress	Stress-responsive signaling pathways

266901_at	Upregulated	May contribute to cellular protection against salt-induced damage	Oxidative stress defense and redox regulation
254396_at	Upregulated	Potential role in maintaining cellular stability under saline conditions	Abiotic stress adaptation and metabolic regulation

As summarized in Table 3.3, all four prioritized genes exhibited consistent upregulation under salinity stress, indicating their potential involvement in key adaptive responses. These genes are functionally associated with osmotic regulation, ion homeostasis, stress-responsive signaling, oxidative stress defense, and metabolic stability, which are critical mechanisms underlying plant salinity tolerance. From a biological perspective, the results suggest that these genes may play regulatory and protective roles in maintaining cellular function under high-salinity conditions. Computationally, the findings demonstrate that integrating transcriptomic feature selection with deep learning-based prioritization can effectively identify biologically relevant candidate genes associated with stress adaptation. Although the ANN model achieved the highest classification performance, the relatively small dataset highlights the need for further validation using larger and independent transcriptomic datasets. Overall, this study addresses a significant gap in salinity-tolerance gene discovery and provides a strong foundation for future multi-omics integration and functional validation studies.

4. REFERENCES

[1] Aziz, M. A., & Masmoudi, K. “Insights Into the Transcriptomics of Crop Wild Relatives to Unravel the Salinity Stress Adaptive Mechanisms”, Int. J. Mol. Sci., vol. 24, no. 12, 2023.

[2] Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. “A Survey of Best Practices for RNA-Seq Data Analysis”, Genome Biology, vol. 17, no. 13, 2016.

[3] Eraslan, G., Avsec, Ž., Gagneur, J. *et al.* “Deep learning: new computational modelling techniques for genomics”, Nat Rev Genet, **20**, 389–403, 2019.

[4] Frontiers Plant Science, "Comparative Transcriptome Analysis Reveals the Adaptive Mechanisms of Halophyte Suaeda dendroides Encountering High Saline Environment," Front. Plant Sci., vol. 15, 2024.

[5] John C, Sahoo J, Sajan IK, Madhavan M, Mathew OK., “CNN-BLSTM based deep learning framework for eukaryotic kinome classification: An explainability based approach”, Comput Biol Chem. 2024.

[6] Min, S., Lee, B., & Yoon, S. “Deep Learning in Bioinformatics”, Briefings in Bioinformatics, vol. 18, no. 5, pp. 851–869, 2017.

[7] S. Toscano, D. Romano, & A. Ferrante, "Molecular Responses of Vegetable, Ornamental Crops, and Model Plants to Salinity Stress," Int. J. Mol. Sci., vol. 24, no. 4, p. 3190, 2023.

Biographies

Nikitha V received the Master's degree in Bioinformatics and Data Science and is currently pursuing a Ph.D. in Computer Science at the Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India. Her research interests include bioinformatics, deep learning, machine learning, transcriptomic data analysis, computational biology, and artificial intelligence applications in agriculture. She has authored and presented research papers in national and international conferences and is actively involved in interdisciplinary research focusing on genomic data analytics and sustainable agricultural technologies.

Sree Kala T is a Professor in the Department of Advanced Computing and Analytics, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India. She has published numerous research articles in reputed journals and conferences and has extensive experience in guiding research scholars in advanced computing, data science, and interdisciplinary research domains.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Modeling and Detection of FoMO-Inducing Content in Social Media Advertising Using Machine Learning Techniques

Sandhya Sabu[1] , A. Ambeth Raja [2*] , V.Devi[3] ,S.Lavanya[4]
[1 & 2 & 3 & 4] PG & Research Department of Computer Science
Thiruthangal Nadar College, Chennai.
[1]itzmesabu@gmail.com ,[2]arajacs1983@gmail.com,
[3]drdevi.venkatesan@gmail.com ,[4]lavanya.tnc20@gmail.com

Abstract

Fear of Missing Out (FoMO) has emerged as a critical psychological driver influencing user engagement with digital advertisements, particularly for lifestyle products. This study proposes FoMO-ML, a machine learning–based framework for the automatic detection of FoMO-inducing patterns in social media advertisements and the evaluation of their behavioral impact. Using the Facebook Ads Dataset, the framework preprocesses textual content, extracts TF–IDF–based feature representations, and applies multiple supervised learning algorithms—including Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting—for classification into FoMO-oriented and non-FoMO categories. Feature importance analysis identifies key linguistic cues such as urgency, scarcity, exclusivity, and social proof, which are strongly associated with heightened user engagement. Behavioral analysis demonstrates that FoMO-classified advertisements consistently elicit higher click-through rates, likes, shares, and comments compared to non-FoMO ads, confirming the practical relevance of the detected patterns. Among the models, Gradient Boosting achieves the highest classification performance, highlighting the advantage of ensemble methods in capturing complex feature interactions. The proposed framework provides a scalable, interpretable, and data-driven approach for understanding and quantifying the influence of FoMO in digital advertising, offering actionable insights for marketers and social media platforms while bridging the gap between computational detection and behavioral outcomes.

Keywords: Fear of Missing Out, Machine learning, social media, Gradient Boosting, Random Forest, Detection, Automation

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. Introduction The rapid expansion of social media platforms has fundamentally transformed the landscape of digital advertising, enabling organizations to reach highly targeted audiences at unprecedented scale and speed. Platforms such as Facebook have become central to modern marketing ecosystems, offering advertisers sophisticated tools to deliver personalized content based on user demographics, preferences, and online behavior. While this data-driven personalization has enhanced advertising efficiency, it has also intensified the use of psychological persuasion techniques designed to influence user decision-making [1-4]. FoMO refers to the anxiety or apprehension that individuals experience when they believe they may miss rewarding experiences, opportunities, or benefits that others are accessing [5-8]. As a result, understanding and detecting FoMO-inducing patterns has become an important research challenge at the intersection of digital marketing, consumer psychology, and computational analytics.

2.Proposed Work

Based on the identified research gaps in existing studies on Fear of Missing Out (FoMO), this work proposes a data-driven, machine learning–based framework for the automated detection of FoMO-inducing patterns in social media advertisements and the assessment of their behavioral impactSuch approaches, although theoretically rich, lack scalability and fail to capture how FoMO is practically embedded within real-world advertising content. The proposed work addresses this limitation by operationalizing FoMO as a detectable linguistic and contextual phenomenon within digital advertisements.

3. Proposed Methodology

The proposed work introduces the FoMO-ML Framework, a structured machine learning pipeline designed to automatically identify FoMO-inducing advertising patterns and analyze their behavioral implications. The methodology follows a sequential and modular process consisting of text preprocessing, feature extraction, supervised classification, and behavioral impact analysis. This design ensures scalability, reproducibility, and adaptability to evolving advertising language.

Feature extraction is performed using Term Frequency–Inverse Document Frequency (TF–IDF) representations, incorporating both unigram and n-gram features to capture isolated urgency terms as well as phrase-level FoMO patterns. These features form the numerical input to supervised learning models.

The proposed work employs a diverse set of supervised learning algorithms to capture different aspects of FoMO-related patterns:

Supervised Learning Algorithms for FoMO Detection

To effectively identify Fear of Missing Out (FoMO)–inducing patterns in social media advertisements, the proposed FoMO-ML Framework leverages a diverse set of supervised learning algorithms. Each algorithm contributes distinct strengths in capturing linear and non-linear relationships, handling high-dimensional feature spaces, and optimizing prediction accuracy. The choice of multiple algorithms also enables comparative analysis and ensures robustness of the detection framework.

1. Logistic Regression (LR)

The probability of a FoMO label $\hat{y}_i = 1$ is given by the sigmoid function:

$$P(y_i = 1 \mid x_i) = \frac{1}{1+e^{-(w^\top x_i+b)}}$$

where $w \in R^M$ is the weight vector and b is the bias term. The model parameters are optimized by minimizing the binary cross-entropy loss:

$$L(w,b) = -\frac{1}{N}\sum_{y=1}^{N}[y_i log P(y_i = 1 \mid x_i) + (1-y_i) log(1 - P(y_i = 1 \mid x_i))]$$

LR enables identification of key textual features that significantly contribute to FoMO classification, such as urgency indicators (“limited time,” “hurry”) and social proof phrases (“trending now”).

2. Support Vector Machine (SVM)

The SVM classifier seeks to identify an optimal hyperplane that maximizes the margin between FoMO and non-FoMO classes. Formally, SVM solves the optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} \parallel w \parallel^2 + C\sum_{i=1}^{N} \xi_i$$

subject to: $y_i(w^\top x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$

where ξ_i are slack variables allowing soft-margin classification and C is a regularization parameter balancing margin maximization with classification error.

3. Random Forest (RF)

Random Forest is an ensemble learning algorithm that builds multiple decision trees using bootstrapped samples and averages their predictions. For each tree $h_k(x_i)$, a subset of features is randomly selected at each split to reduce correlation between trees. The final prediction is obtained via majority voting: $\hat{y}_i = mode\{h_1(x_i), h_2(x_i), \dots, h_K(x_i)\}$

where K is the number of trees in the forest. RF captures non-linear relationships and complex feature interactions, which are common in persuasive advertisement language

4. Gradient Boosting Models (GBM)

Formally, for iteration mmm, the model is updated as: $F_m(x_i) = F_{m-1}(x_i) + \gamma_m h_m(x_i)$

where $h_m(x_i)$ is the mth weak learner, γ_m is a learning rate parameter, and $F_0(x_i)$ is initialized as the mean of the target labels. The model minimizes a differentiable loss function, such as logistic loss for binary classification.

Expected Contribution of the Proposed Work

The proposed FoMO-ML Framework transforms FoMO from a latent psychological construct into a **computationally observable and measurable phenomenon**. By leveraging real-world advertising data, machine learning techniques, and engagement analytics, the framework addresses key limitations in existing FoMO research related to scalability, objectivity, and practical relevance. The proposed work contributes to digital marketing analytics, consumer behavior research, and responsible advertising practices by providing an automated and interpretable approach to identifying psychologically persuasive advertising strategies on social media platforms.

4. Results

Feature Importance Analysis

Table 1 presents the top ten features contributing to FoMO detection, derived from the Random Forest model's feature importance scores. This highlights the linguistic cues most predictive of FoMO orientation, such as urgency, scarcity, exclusivity, and social proof, demonstrating the multi-faceted nature of persuasive advertising language

Table 1: Top 10 Features Contributing to FoMO Detection (Random Forest)

Rank	Feature	Importance Score
1	ending soon	0.082
2	limited offer	0.079
3	trending now	0.074
4	members only	0.068
5	only today	0.065
6	best seller	0.061
7	few left	0.058
8	exclusive deal	0.056
9	top rated	0.053
10	hot item	0.051

Comparative Algorithm Analysis Across Engagement Metrics

The relationship between classifier predictions and engagement metrics was also analyzed by stratifying ads based on predicted FoMO scores. **Table 2** presents Gradient Boosting model predictions across different engagement ranges.

Table 2. Predicted Probability Range

Predicted Probability Range	Avg CTR (%)	Avg Likes	Avg Shares	Avg Comments
0.0 – 0.2	3.8	61	22	10
0.2 – 0.4	5.1	74	27	12
0.4 – 0.6	6.5	87	33	15
0.6 – 0.8	8.2	105	42	19
0.8 – 1.0	9.6	121	48	23

5. Conclusion

This study presents the FoMO-ML Framework, a novel machine learning–based approach for detecting Fear of Missing Out (FoMO) patterns in social media advertisements and evaluating their behavioral impact on user

engagement. By leveraging the Facebook Ads Dataset and employing a combination of supervised learning algorithms—including Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting—the framework demonstrates robust performance in classifying advertisements into FoMO-oriented and non-FoMO categories. The results indicate that ensemble models, particularly Gradient Boosting, outperform simpler models by capturing subtle, multi-feature interactions, achieving an accuracy of 90.3%, F1-Score of 90.3%, and an AUC of 0.94. Feature importance analysis further confirms that linguistic patterns related to urgency and social validation are the strongest predictors of FoMO-oriented content. Behavioral analysis demonstrates a clear and statistically significant increase in user engagement metrics, including click-through rates, likes, shares, and comments, when exposed to FoMO-inducing advertisements.

References

[1]. Abel, J. P., Buff, C. L., & Burr, S. A. (2016). Social media and the fear of missing out: Scale development and assessment. *Journal of Business and Economics Research, 14*(1), 33–44. https://doi.org/10.19030/jber.v14i1.9554

[2]. Alt, D., & Boniel-Nissim, M. (2018). Parent-adolescent communication and problematic internet use: The mediating role of fear of missing out (FoMO). *Journal of Family Issues, 39*(13), 3391–3409. https://doi.org/10.1177/0192513X18783493

[3]. Ansari, S., Ansari, G., Ghori, M., & Kazi, A. G. (2019). The impact of brand awareness and social media content marketing on consumer purchase decision. *Journal of Public Value and Administration Insights, 2*, 5–10. https://doi.org/10.31580/jpvai.v2i2.896

[4]. Aral, S. (2011). Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science, 30*(2), 217–223. https://doi.org/10.1287/mksc.1100.0596

[5]. Ariel, Y., & Avidar, R. (2015). Information, interactivity, and social media. *Atlantic Journal of Communication, 23*(1), 19–30. https://doi.org/10.1080/15456870.2015.972404

[6]. Enginkaya, E., & Sağlam, M. H. (2025). Social commerce platforms role in FoMO driven consumption: The hidden psychological toll of continuous use. *IEEE Access, 13*, 77395–77417. https://doi.org/10.1109/ACCESS.2025.3565127

[7]. Elhai, J. C., Levine, R. D., Dvorak, B. J., & Hall, B. J. (2016). Fear of missing out, need for touch, anxiety and depression are related to

problematic smartphone use. *Computers in Human Behavior, 63*, 509–516. https://doi.org/10.1016/j.chb.2016.05.079

[8]. Hattingh, M., Dhir, A., Ractham, P., Ferraris, A., & Yahiaoui, D. (2022). Factors mediating social media-induced fear of missing out (FoMO) and social media fatigue: A comparative study among Instagram and Snapchat users. *Technological Forecasting and Social Change, 185*, 122099. https://doi.org/10.1016/j.techfore.2022.122099

Biographies



Mrs. Sandhya Sabu received the bachelor's degree in Computer Science from University of Madras in 2013, the master's degree in Computer Science & its Applications from IGNOU in 2018, and pursuing as Research Scholar in Computer Science from Thiruthangal Nadar College(aff.. to University of Madras) from 2026, respectively. She is currently working as an Assistant Professor at the Department of Computer Science in Thiruthangal Nadar College, Chennai. Her research areas include Artificial Intelligence and Machine Learning.



Dr. A. Ambeth Raja, M.Sc., M.Phil., Ph.D., is an Associate Professor and Head of the PG & Research Department of Computer Science at Thiruthangal Nadar College with over 18 years of teaching experience. He specializes in Design and Analysis of Algorithms and Data Mining Applications and has presented and published more than 25 papers, including an international presentation in Malaysia. He has served as a resource person, reviewer, and editorial board member for several journals and conferences. His achievements have earned him notable honours such as the International Distinguished Researcher Award 2020 and the Young Scholar Award 2020. He has also authored books on Computer Organization, C++ Programming, Artificial Intelligence, and Machine Learning.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

The Next Generation of AI for Social Transformation: Opportunities and Governance Challenges

Deepti Chhabra, Research Scholar, School of Law, UPES

Email- Deepti.29708@stu.upes.ac.in orcid- 0009-0006-6172-3423

Gagandeep Kaur, Senior Associate Professor, School of Law, UPES

Email- gkaur@ddn.upes.ac.in orcid- 0000-0003-3657-2134

Abstract. Artificial intelligence is changing every industry, and it is considered an important tool for sustainable development. This article investigates the topic of "AI for Social Impact" (AI4SI): the use of advanced AI technologies, such as foundation models, generative AI, cloud computing, and data-centered approaches, to provide services in such areas as public sector, health care, education, environment conservation, disaster response, and poverty reduction. Based on the narrative literature review of the AI4SI and responsible AI, the article reveals the factors of governance, which prevent people from trusting AI: algorithmic bias, insufficient transparency and accountability, privacy concerns, lack of data governance, digital inequality, and fragmentation of regulation. Moreover, the article describes the conditions under which implementation of AI4SI initiatives becomes successful, including local approach, multidisciplinary collaboration, involvement of stakeholders, and impact assessment, and provides a short list of governance readiness criteria to be used by practitioners and reviewers in evaluating AI4SI projects.

Keywords: AI for social Impact; social transformation; AI governance; Responsible AI; SDGs; Digital Inclusion.

1. INTRODUCTION

What began as an esoteric discipline is now a tool that can radically transform the way we structure our societies, economies, and governance systems. Foundation models, generative AI, large language models, and multi-agent systems are just some of the recent developments that have enabled applications to expand and open new possibilities for solving social challenges. This evolution has helped shape "AI for Social Impact" (AI4SI), a branch of AI research distinct from traditional AI research in that it puts emphasis on societal impact for disadvantaged communities as opposed to novelty and performance improvements [5]. AI4SI focuses on topics such as poverty reduction, public health, education, agriculture, disaster response, and sustainability [1,5] and requires the collaboration of multiple disciplines [8].

Previous work has typically conflated "AI for Social Good" and AI4SI; this paper, like recent work in the field, will consider AI4SI to be the more accurate label for interventions that have a specific social outcome and a defined beneficiary group [1,5,6]. It will use "social transformation" to describe the larger, more long-term societal transformation that a set of AI4SI interventions can help achieve. The 2016 White House Workshop on AI for Social Good is commonly acknowledged as the event that marked the consolidation of the field [2].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

While these opportunities exist, application of AI to social sectors has been associated with continuous governance challenges such as algorithmic bias, poor transparency and accountability, privacy and security concerns, fake news and digital divide [9]. Events in 2025-2026, such as the emergence of socially impactful systems developed using foundation models, have further brought up the issue of human control, fairness, and compliance [6]. The aim of this paper is to explore the opportunities presented by next generation AI in social change, its governance challenges, and the operational checklist thereof.

2. REVIEW METHODOLOGY

This paper is not a systematic review or an empirical study but rather a narrative one. The sources used come from three sources: (i) basic and modern studies on AI4SI (e.g., [5,6,8,11,13]); (ii) responsible-AI and governance studies (e.g., [7,9,18]); and (iii) applications of AI4SI in computer science and public health studies (e.g., [14,15,16,17]) along with additional policy studies like the United Nations SDGs [1] and the most recent public opinion surveys [10]. These sources have been identified using the references of surveys and conference tracks of established surveys and conference tracks of AI4SI (AAAI, IJCAI, AAMAS). It does not purport to be comprehensive but is a way of synthesizing common themes and using actual deployment experience to make the governance checklist presented in Section 4 meaningful.

3. CASE STUDIES IN AI4SI DEPLOYMENT

In order to discuss the issue of governance based on practice rather than principles only, this paper analyses three applications of AI4SI that have been described in the literature. The governance lesson that has been learned from each application is presented in Table 1 below.

Case	Deployment	Governance lesson
Maternal heath engagement, India [8]	Call prioritization for an NGO (ARMMAN) for mothers who are likely to disengage from an automated health line, assessed using a real randomized deployment.	Effective when the deploying institution retains interpretable oversight and override power over the decisions made by algorithms.
Humanitarian aid targeting, Togo [15]	Mobile phone data ML prioritization for COVID-19 emergency cash transfer without survey data.	Increases coverage in low data environments but necessitates clear and legal limits to data usage and consent.
Flood forecasting, multi-country [16]	Validation of ML flood warning system based on historical gauge data and warning dissemination.	Protective efficacy is contingent on ongoing model validation and a local partner institution to implement outcomes.

Table 1. Selected AI4SI deployment and governance lessons.

Common threads across these examples include three factors irrespective of sector: a named institution responsible for algorithmic decision-making, an explicit limitation to data use and consent, and a local partner capable of maintaining the system post initial technology implementation. It is upon these factors, rather than assumptions, that the checklist in Section 5 is based.

4. CURRENT LANDSCAPE AND EMERGING TRENDS

Aside from the above-discussed cases, the development of AI4SI technology is becoming more interdisciplinary in recent times. The collaboration of computer scientists, policymakers, doctors, social workers, environmental scientists, and community members

will help ensure that the solution is contextually relevant and socially acceptable [8]. In this regard, stakeholder participation throughout the process of design, deployment, and evaluation of the AI4SI project represents application of the rights-based approach to disability with respect to technical governance ("nothing about us without us") [7]. There are two trends worth mentioning at the field level: the use of foundation models decreases the cost of creation of task-specific systems where technical knowledge and data are limited [6], and cloud infrastructure reduces the cost of implementing AI4SI solutions in the public sector. Public discussion about AI is developing amidst worries about misinformation, unemployment, discrimination, and lack of regulation [10], which shifts the attention of the field from technical aspects to those that were discussed directly in Section 3.

5. TOWARDS OPERATIONAL GOVERNANCE : A PRACTITIONER CHECKLIST

Criteria like fairness, transparency, and accountability are discussed in many papers in the AI4SI literature [9,18]; however, very few papers convert these criteria to a format that can be used in practice. Drawing on the lessons learned from deployment (Section 3) [8,15,16] and the discussion of design factors in [9], we present a six-dimension governance readiness checklist in Table 2 that should be used in the design and pre-deployment phases of a project rather than as a certification criterion. In each dimension, we combine an issue related to governance with a question, which makes it clear what projects will have high risk on a certain dimension because they cannot say "yes" to the question or have a mitigation strategy.

Dimensions	**Pre – deployement questions**
Fairness & Bias	Have there been tests of the representativeness of training and deployment data on the target population?
Transparency	Can affected users and oversight organizations receive an easily understandable account of how decisions are made by the system?
Accountability	Is there someone accountable for results, as well as for resolving any issues?
Privacy & Data Governance	Does data gathering, management, and transfer have clear legal policy in place?
Inclusion & Participation	Have the affected communities been consulted in designing and evaluating the system as well as deploying it?
Sustainability	Has there been provision for running the system once external technical assistance has stopped?

Table 2. AI4SI governance – readiness checklist.

6. THE FUTURE OF AI FOR SOCIAL TRANSFORMATION

The development path for socially transformative AI is contingent upon the interplay between technological advancement and the institutions within which it operates. Three issues consistently arise from the literature surveyed. Firstly, AI for SI applications must be developed with regard for the context in which they operate and be the product of participatory design that engages the stakeholders affected by its implementation [8]. Secondly, scalability and sustainability continue to be problems, in that while many AI for SI projects are successful early on, few continue to have an impact in the long term, and

local governments, NGOs, and other public bodies require the capacity and skills to maintain the systems without reliance on external technical assistance [5]. Finally, evaluation practice for the systems and their impact must expand past purely technological measures to consider socioeconomic and ethical outcomes based on fieldwork and stakeholder assessment [7].

Making the tools of AI4SI more widely available to experts in the field, civil society organizations, and communities, in conjunction with greater AI literacy among policymakers and other stakeholders, will likely foster both inclusive innovation and more resilient citizen participation. The development of foundation models is anticipated to make the construction of tools even more efficient and lower the cost through reducing the task-specificity of the systems [6], but in doing so, such generalization also brings about risk – communities that are under-represented within the training datasets become disproportionately vulnerable to bias and exclusion, while insufficient transparency and context within foundation models makes the development of robust and equitable AI systems difficult. As the technology becomes increasingly adopted, it is likely that research will move from technical deployment to governance and equitable distribution of AI4SI benefits.

7. CONCLUSION

Social transformations through AI is one of the most impactful use cases of artificial intelligence with proven potential in the areas of health, education, sustainable environment, poverty reduction, disaster response, public services as shown by the use cases of Section 3. Unlocking the potential will require addressing several interlinked challenges including: socially contextualized, sustainable systems; representative data; and evaluation methodology that measures impact on social outcome rather than proving the capability. Social transformations through AI will also depend on ongoing interdisciplinary collaboration, wider AI literacy of the policy makers, practitioners and citizens, and engagement of stakeholders throughout the life cycle so that the transparency and inclusion are designed in. Therefore, the impact that AI4SI can have on society will be not only dependent on the technology but the integration of it in socio-cultural and institutional system. Governance checklist provided in Section 5 based on the case studies of Section 3 is proposed as an attempt to assess such integration; further validation and refinement of it based on the additional case study should be carried out in the future work. Next generation of AI can become a reliable source of social transformations if all these aspects are pursued in parallel.

8. REFERENCES

[1] United Nations. (2015). Transforming Our World: The 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda.

[2] White House Office of Science and Technology Policy. (2016). Workshop on AI for Social Good. https://cra.org/ccc/events/ai-social-good/.

[3] Association for the Advancement of Artificial Intelligence. (2024). Call for the Special Track on AI for Social Impact, AAAI-24. https://aaai.org/aaai-24-conference/call-for-the-special-track-on-ai-for-social-impact/.

[4] International Joint Conferences on Artificial Intelligence. (2025). Call for Papers and Projects: Multi-Year Track on AI and Social Good. https://2025.ijcai.org/call-for-papers-and-projects-multi-year-track-on-ai-and-social-good-special-track/.

[5] Shi, Z., Wang, C., & Fang, F. (2020). AI for social impact: A survey. arXiv. https://arxiv.org/abs/2001.01818.

[6] Zhao, Y., Boehmer, N., Taneja, A., & Tambe, M. (2025). Towards foundation-model-based multi-agent systems to accelerate AI for social impact. Proceedings of AAMAS 2025, 2901–2907.

[7] Charlton, J. I. (1998). Nothing About Us Without Us: Disability Oppression and Empowerment. University of California Press.

[8] Mate, A., Madaan, L., Taneja, A., Madhiwalla, N., Verma, S., Singh, G., Hegde, A., Varakantham, P., & Tambe, M. (2022). Field study in deploying restless multi-armed bandits: Assisting non-profits in improving maternal and child health. Proceedings of the AAAI Conference on Artificial Intelligence, 36(11), 12017–12025. https://doi.org/10.1609/aaai.v36i11.21460.

[9] Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2021). How to design AI for social good: Seven essential factors. In Ethics, Governance, and Policies in Artificial Intelligence (pp. 125–151). Springer. https://doi.org/10.1007/978-3-030-81907-1_9.

[10] Pew Research Center. (2025). How the U.S. Public and AI Experts View Artificial Intelligence. https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/.

[11] Tambe, M., & Rice, E. (Eds.). (2018). Artificial Intelligence and Social Work. Cambridge University Press. https://doi.org/10.1017/9781108669016.

[12] Gomes, C., Dietterich, T., Barrett, C., Conrad, J., Dilkina, B., Ermon, S., Fang, F., Farnsworth, A., Fern, A., Fern, X., & Fink, D. (2019). Computational sustainability: Computing for a better world and a sustainable future. Communications of the ACM, 62(9), 56–65. https://doi.org/10.1145/3339399.

[13] Tambe, M., Fang, F., & Wilder, B. (Eds.). (2022). AI for Social Impact. https://ai4sibook.org/.

[14] Yadav, A., Chan, H., Jiang, A. X., Xu, H., Rice, E., & Tambe, M. (2016). Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. Proceedings of AAMAS 2016, 740–748.

[15] Aiken, E., Bellue, S., Karlan, D., Udry, C., & Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid. Nature, 603, 864–870. https://doi.org/10.1038/s41586-022-04484-9.

[16] Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., & Begelman, G. (2022). Flood forecasting with machine learning models in an operational framework. Hydrology and Earth System Sciences, 26(15), 4013–4032. https://doi.org/10.5194/hess-26-4013-2022.

[17] Shi, Z. R., Lizarondo, L., & Fang, F. (2021). A recommender system for crowdsourcing food rescue platforms. Proceedings of the Web Conference 2021, 857–865. https://doi.org/10.1145/3442381.3449787.

[18] Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K. S., Myles, P., & Grainger, D. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ, 368. https://doi.org/10.1136/bmj.l6927.

Biographies

Deepti Chhabra
Doctoral Scholar, UPES, Dehradun



Deepti Chhabra is a Doctoral Scholar at University of Petroleum & Energy Studies, Dehradun. She holds an LL.M. with specialization in Cyber Security and Digital Laws, wherein she was recognized as a merit holder and received formal academic appreciation. She completed her B.A. LL.B. (Hons.) with specialization in Criminal Laws. Her academic and professional engagements include extensive work in legal research, contract lifecycle management, drafting, and dispute resolution across diverse legal forums. She has authored several scholarly publications in reputed journals and has actively participated in national moot court competitions. Her research interests lie at the intersection of technology law, cyber regulation, and criminal justice.

Dr Gagandeep Kaur

Senior Associate Professor in Law, UPES, Dehradun



Dr. Gagandeep Kaur is LL.B. (Gold Medalist), LL.M. (Gold Medalist) & Ph.D. in Law with distinguished expertise in Cyber Laws, Information Technology Laws, Cyber Crimes & Data protection Laws. She has more than 16 years of teaching and research experience in prestigious Universities and is presently associated with the School of Law, UPES Dehradun as Senior Associate Professor in Law. She has authored and edited more than 25 books. She has published over 160 articles in peer-reviewed, UGC-Care, SCOPUS Indexed & Web of Science Journals. She has published several book chapters with Springer, Wiley, Palgrave Macmillan, Apple Academic Press & CRC Press. She has presented over 160 research papers at national and international conferences.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

DSTAQ-Net: A Dynamic Spatio-Temporal Deep Learning Framework for Multi-Pollutant Air Quality Prediction Using Distributed Sensor Networks

M.Divya[1] V. Devi[2]* **D.Ravikumar[3]**
[1]Research Scholar [2]Professor [3] Associate Professor,
[1&2] PG & Research Department of Computer Science
[3]Department of Medical Electronics,
[1&2]Thiruthangal Nadar College, Chennai.
[3]SRM Valliammai Engineering College, Chennai
[1]mdivyamca90@gmail.com[2]drdevi.venkatesan@gmail.com[3]
ravikumar.dinakaran@gmail.com

Abstract

Accurate and timely air quality prediction is critical for mitigating the adverse impacts of air pollution on public health, urban sustainability, and environmental governance. However, existing forecasting models often struggle to capture the complex nonlinear interactions and dynamic spatio-temporal dependencies inherent in large-scale air quality sensor networks, particularly under noisy and heterogeneous measurement conditions. To address these challenges, this paper proposes a deep learning–based spatio-temporal framework, named DSTAQ-Net, for real-time and multi-horizon prediction of air pollutant concentrations using distributed IoT sensor networks. The proposed framework integrates sensor reliability assessment, dynamic spatial graph construction, spatio-temporal feature encoding, attention-based learning, and multi-pollutant forecasting into a unified architecture. A dynamic graph convolutional network is employed to model evolving spatial relationships among monitoring stations based on geographic proximity, pollutant correlation, and meteorological similarity, while gated recurrent units capture temporal dependencies in pollutant dynamics. An attention mechanism further enhances the model by adaptively weighting the most influential sensors, pollutants, and historical time steps, enabling robust learning under sensor noise and data sparsity. The forecasting layer jointly predicts multiple air quality indicators, including PM2.5, PM10, NO_2, CO, O_3, and AQI, across short-term and long-term horizons. The framework is evaluated using two public benchmark datasets, namely the Beijing Multi-Site Air Quality Dataset and the UCI Air Quality Sensor Dataset. Experimental results demonstrate that DSTAQ-Net achieves significantly lower prediction error and higher stability compared with state-of-the-art spatio-temporal deep learning models. These findings confirm the effectiveness of the proposed approach for reliable air quality forecasting and its suitability for deployment in smart city and environmental monitoring applications.

Keywords: Machine learning, Prediction, Air quality, temporal features

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Introduction

Air pollution has emerged as one of the most critical global environmental and public health challenges of the twenty-first century, with rapid urbanization, industrial expansion, vehicular emissions, and climate variability contributing to a continuous deterioration of atmospheric quality in both developed and developing regions. According to international health and environmental agencies, prolonged exposure to fine particulate matter and toxic gaseous pollutants such as PM2.5, PM10, NO_2, SO_2, CO, and O_3 significantly increases the risk of respiratory diseases, cardiovascular disorders, premature mortality, and ecological degradation. In densely populated urban environments, pollution episodes often occur unpredictably and can escalate rapidly due to meteorological changes, traffic congestion, industrial activity, and human behavior, making real-time monitoring and accurate forecasting a crucial requirement for public safety, environmental governance, and smart city management [1-3].

Air pollution is governed by nonlinear physical, chemical, and meteorological processes, including atmospheric diffusion, chemical reactions, wind flow, humidity, temperature inversions, and urban topology, which interact across both space and time. Consequently, simple statistical models such as autoregressive integrated moving average (ARIMA), linear regression, and basic interpolation techniques fail to capture the intricate spatio-temporal dependencies and nonlinear correlations that characterize real-world air quality dynamics [4-7]. Even conventional machine learning approaches, including support vector machines, decision trees, and random forests, while capable of modeling certain nonlinearities, typically treat spatial and temporal dimensions independently and struggle to generalize under highly dynamic environmental conditions [8-13].

The framework is designed to process multi-pollutant time-series data collected from distributed monitoring nodes, transform them into structured spatio-temporal representations, and generate accurate short-term and long-term forecasts for key air quality indicators, including particulate matter and gaseous pollutants. Unlike traditional models that require extensive domain-specific feature engineering and calibration, the proposed deep learning approach learns predictive patterns directly from data, enabling adaptive modeling across different geographic regions, seasonal variations, and emission patterns. Moreover, the use of sensor networks allows the system to operate at high spatial resolution, making it suitable for fine-grained urban pollution mapping, localized alert generation, and micro-environmental analysis.

Research Gaps

Although existing studies demonstrate that deep learning-based spatio-temporal models significantly improve air quality prediction, several critical gaps remain. First, most current approaches rely on static or semi-static spatial relationships among monitoring stations, whereas real pollution transport is highly dynamic and influenced by meteorology, urban morphology, and emission patterns, requiring more adaptive and context-aware spatial modeling. Second, many deep architectures are computationally intensive and difficult to deploy in real-time IoT-based sensor networks, limiting their practical applicability in smart city environments. These gaps motivate the need for a lightweight yet powerful spatio-temporal deep learning framework that is robust, adaptive, and deployable within real-world sensor network infrastructures for reliable air quality and pollution level prediction.

Proposed Work

Existing air quality prediction models predominantly rely on static spatial correlations, heavy deep learning architectures, and centralized computation, which limits their effectiveness in real-world IoT-based sensor networks characterized by noisy data, heterogeneous sampling rates, and dynamic environmental conditions. Moreover, many models are computationally expensive and unsuitable for real-time deployment, and few incorporate physical atmospheric influences such as meteorology and pollutant diffusion patterns. To address these limitations, this work proposes a lightweight, adaptive, and physically informed deep learning framework that integrates dynamic spatio-temporal modeling, sensor reliability awareness, and attention-driven feature selection to enable accurate and real-time air pollution forecasting over large-scale sensor networks.

Overview of the Proposed Framework

The proposed system, named DSTAQ-Net (Dynamic Spatio-Temporal Air Quality Network), is designed to perform real-time and multi-horizon prediction of air quality indicators using data from distributed IoT-based sensor networks. It consists of five interconnected modules:

1. Data Acquisition and Sensor Reliability Layer
2. Dynamic Spatial Graph Construction
3. Spatio-Temporal Feature Encoding
4. Attention-based Prediction Network
5. Multi-Pollutant Forecasting Layer

The framework jointly models:

- Spatial relationships between sensors using a dynamic graph structure,
- Temporal dependencies using gated recurrent networks,
- Pollutant interactions using attention-based fusion.

Data Preprocessing

To ensure the reliability, consistency, and learning efficiency of the proposed deep spatio-temporal framework, a structured data preprocessing pipeline is applied to all sensor measurements prior to model training and evaluation. Given that air quality sensor networks often suffer from missing values, noise, and heterogeneous scales, multiple preprocessing operations are performed to transform the raw data into high-quality spatio-temporal representations.

Training Algorithm

The model is trained using Adam optimizer with Huber loss to handle sensor noise.

Input: Sensor data X, meteorology M, graph G

Output: Predicted pollution levels Y
1. Normalize X and M
2. Construct dynamic adjacency matrix A
3. Apply GCN to X using A
4. Feed GCN output into GRU

5. Apply attention to hidden states
6. Predict future pollutant values
7. Compute loss
8. Update weights

Repeat until convergence

The Dynamic Spatial Graph Construction layer models the physical and functional relationships among sensor nodes. Unlike traditional approaches that use static adjacency matrices based solely on geographic proximity, the proposed framework constructs a dynamic graph that evolves over time based on pollution correlation, meteorological similarity, and spatial distance. Spatio-Temporal Feature Encoding is performed by combining graph-based spatial learning with sequence-based temporal modeling. In the spatial dimension, graph convolutional networks are applied over the dynamically constructed sensor graph to aggregate information from neighboring nodes in a weighted manner. The final Multi-Pollutant Forecasting Layer generates future predictions for multiple air quality indicators in a unified manner. Instead of modeling each pollutant independently, the framework performs joint multi-output regression, capturing the interdependencies among different pollutants.

Experimental Evaluation

Experiments were conducted using two publicly available benchmark datasets to validate the effectiveness and generalizability of the proposed DSTAQ-Net framework. Dataset 1 was collected by the Beijing Municipal Environmental Monitoring Center and contains hourly air quality and meteorological measurements from 36 monitoring stations between 2013 and 2017. It includes PM2.5, PM10, SO_2, NO_2, CO, O_3, temperature, humidity, wind speed, wind direction, and pressure.

Performance on Noisy Sensor Data

To evaluate robustness under real IoT conditions, the model was tested on the UCI Air Quality dataset.

Table 1. Performance on UCI Air Quality Dataset

Parameter	MAE	RMSE	MAPE (%)	R^2
CO	0.38	0.51	9.12	0.917
NOx	11.92	16.04	10.36	0.902
NO_2	8.41	12.19	9.48	0.911
Temperature	0.84	1.17	4.26	0.961
Humidity	2.91	4.28	5.87	0.946

These results confirm that DSTAQ-Net remains resilient against sensor drift and noise, making it suitable for real-world deployments.

Comparative Evaluation with Existing Methods

Table 2. Comparative Performance on PM2.5 Prediction

Model	MAE	RMSE	MAPE (%)	R^2
CNN-LSTM (Guo et al., 2023)	9.61	13.42	11.38	0.901
ST-Attention (Chen et al., 2025)	8.43	11.96	9.82	0.921
Residual ST-LSTM (Li et al., 2024)	7.98	11.14	9.21	0.934
Proposed DSTAQ-Net	7.14	10.21	8.92	0.943

The proposed framework consistently outperforms all baseline models, demonstrating superior accuracy and generalization due to its dynamic spatial modeling and attention-based learning.

Conclusion

This study presented a deep learning–based spatio-temporal framework, termed DSTAQ-Net, for accurate and reliable air quality and pollution level prediction using distributed sensor networks. The proposed framework was designed to address key limitations of existing approaches, including static spatial modeling, sensitivity to sensor noise, limited adaptability to dynamic environmental conditions, and lack of scalability for real-time smart city deployments.Comparative analysis with recent state-of-the-art spatio-temporal deep learning models from the literature showed that DSTAQ-Net outperforms existing CNN–LSTM, attention-based, and residual deep learning architectures, particularly in PM2.5 forecasting accuracy. This performance improvement can be attributed to the dynamic spatial graph modeling that adapts to changing pollutant transport patterns and the attention mechanism that selectively emphasizes the most influential sensors, pollutants, and time steps.

References

[1] Chen, G., Chen, S., Li, D., *et al.*: A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention. *Scientific Reports* **15**, 3685 (2025). https://doi.org/10.1038/s41598-025-88086-1

[2] Guo, R., Zhang, Q., Yu, X., Qi, Y., Zhao, B.: A deep spatio-temporal learning network for continuous citywide air quality forecast based on dense monitoring data. *Journal of Cleaner Production* **414**, 137568 (2023).

[3] Agbehadji, I.E., Obagbuwa, I.C.: Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. *Atmosphere* **15**(11), 1352 (2024). https://doi.org/10.3390/atmos15111352

[4] Kheder, A., Foreback, B., Wang, L., Liu, Z.S., Boy, M.: Deep spatio-temporal neural network for air quality reanalysis. In: Petersen, J., Dahl, V.A. (eds.) *Image Analysis (SCIA 2025)*. Lecture Notes in Computer Science, vol. 15725, pp. 71–84. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-95911-0_6

[5] Pak, U., Kim, C., Ryu, U., Sok, K., Pak, S.: A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere & Health* **11**, 883–895 (2018). https://doi.org/10.1007/s11869-018-0585-1

[6] Du, S., Li, T., Yang, Y., Horng, S.J.: Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering* **33**, 2412–2424 (2021). https://doi.org/10.1109/TKDE.2019.2954510
[7] Pichelstorfer, L., *et al.*: Towards automated inclusion of autoxidation chemistry in models: from precursors to atmospheric implications. *Environmental Science: Atmospheres* **4**, 879–896 (2024).
[8]Kalaiselvi, S., Anitha, V., Manimaran, V., *et al.*: Air quality prediction using multi-source remote sensing data integration with hybrid deep learning framework. *Scientific Reports* **15**, Article 32466 (2025). https://doi.org/10.1038/s41598-025-32466-0
[9] Li, D., Liu, J., Zhao, Y.: Prediction of multi-site PM2.5 concentrations in Beijing using CNN-BiLSTM with CBAM. *Atmosphere* **13**(10), 1722 (2022). https://doi.org/10.3390/atmos13101719
[10] Yang, H., Zhu, Z., Li, C., Li, R.: A novel combined forecasting system for air pollutant concentrations based on fuzzy theory and optimization of aggregation weights. *Applied Soft Computing* **86**, 105972 (2020). https://doi.org/10.1016/j.asoc.2019.105972
[11] Foreback, B., *et al.*: A new implementation of FLEXPART with Enviro-HIRLAM meteorological input and a case study during a heavy air pollution event. *Big Earth Data* **8**(2), 397–434 (2024).
[12] Foreback, B., *et al.*: Severe haze episodes in Beijing may be influenced by emissions in far western China. *Air Quality, Atmosphere & Health* (December 2024).
[13] Gao, C., *et al.*: Two-way coupled meteorology and air quality models in Asia: a systematic review and meta-analysis of impacts of aerosol feedbacks on meteorology and air quality. *Atmospheric Chemistry and Physics* **22**(8), 5265–5329 (2022).

Biographies



Ms. M. Divya received her Bachelor of Science (B.Sc.) degree from the University of Madras in 2010, her Master of Computer Applications (MCA) degree from the University of Madras in 2013, and her Master of Philosophy (M.Phil.) degree from St. Peter's University in 2015. She is currently working as an Assistant Professor in the Department of Bachelor of Computer Applications (BCA) and is also a Research Scholar at Thiruthangal Nadar College. Her research interests include Artificial Intelligence, Machine Learning, Deep Learning, Data Mining, Data Warehousing, and Air Quality Prediction using Machine Learning and Deep Learning techniques



Dr. V. Devi, Principal – Thiruthangal Nadar College, Chennai, and Senate Member in University of Madras. She received her undergraduate degree, Master degree and Ph. D from University of Madras. She published 60 research papers/article in peer review indexed Journals, Books and Book chapters in reputed publishers. She has more than 25 years of experience in teaching and research field. She also presented various academic as well as research-based papers at several national and international conferences in South East Asian Countries. Her research area includes Data Science & Analytics, Cloud Computing, Software Engineering, Digital Image Processing, and IoT Systems and Information System Management.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Early Prediction of Heart Failure Progression Using an Attention-Enhanced LSTM Framework

***S.Bhuvana*[1] *V. Devi*[2]* D.*R. Ashwin Kumar*[3]
[1]Research Scholar [2]Professor
[1&2]PG & Research Department of Computer Science
[3]Department of Computer Science (Data Science), Manipal Academy of Higher Education, Bengaluru
[1&2]Thiruthangal Nadar College, Chennai.
[1]bhuvanajey14@gmail.com2drdevi.venkatesan@gmail.com[3]
ashwinkumardr26@gmail.com

Abstract:
Early identification of heart failure progression is critical for improving patient outcomes and enabling timely clinical interventions. This study proposes a predictive deep learning framework for early detection of heart failure progression using structured clinical records, with a focus on modeling temporal dependencies among key physiological and biochemical indicators. The proposed approach employs an embedding-based Long Short-Term Memory (LSTM) architecture augmented with an attention mechanism to effectively capture sequential patterns and feature-level importance within patient data. The model is trained and evaluated using a standardized experimental protocol, and its performance is assessed through widely accepted metrics, including accuracy, precision, recall, F1-score, specificity, and AUC–ROC. Experimental results demonstrate that the proposed model achieves superior predictive performance, attaining an accuracy of 91.24% and an AUC of 0.94 on the test dataset, while maintaining balanced sensitivity and specificity. Comparative analysis with existing machine learning and deep learning models further confirms the effectiveness of the proposed architecture in capturing progression-related patterns that static models fail to identify. The findings highlight the potential of sequential deep learning models as reliable decision-support tools for early heart failure risk stratification and progression monitoring.

Keywords: Prediction; Machine learning; Deep learning; Time series data; Long Short term memory

Introduction

Heart failure (HF) is a chronic, progressive cardiovascular syndrome in which the heart is unable to pump sufficient blood to meet the metabolic demands of the body. It represents one of the most significant global health challenges, affecting millions of individuals worldwide and accounting for substantial morbidity, mortality, and healthcare expenditure. HF is not a static condition; rather, it evolves through a continuum of physiological

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

deterioration marked by periods of clinical stability interspersed with acute episodes of decompensation[1-3].

Traditional machine learning methods such as logistic regression, decision trees, and support vector machines have been applied to HF prediction, but these models typically rely on handcrafted features or static summaries of time-series data, such as mean, variance, or last-observed values. Such representations inevitably discard temporal dependencies and fail to model how changes over time influence future outcomes[4-8].

The primary objective of this study is to develop and validate a deep learning–based framework for the early prediction of heart failure progression using structured clinical records. Specifically, the work proposes an embedding-enhanced LSTM model with an attention mechanism that can differentiate between stable and progressively worsening HF states by learning both temporal dependencies and feature-level importance from sequential clinical data. The model is designed to identify patients at risk of deterioration within predefined future intervals, enabling earlier and more accurate risk stratification compared with conventional machine learning and static predictive approaches.

Proposed Work

This section outlines the detailed design of the proposed predictive framework for early detection of heart failure progression using deep learning. Based on the gaps identified in the literature—specifically the need for temporal modeling of disease progression beyond static classification and ensemble methods—this work develops a Long Short-Term Memory (LSTM)–based multivariate sequence model tailored to the clinical characteristics of heart failure patients.

Problem Formulation

The central objective is to transform this tabular clinical dataset into a temporal prediction task where the goal is to forecast heart failure progression (death_event) based on longitudinal sequences of clinical parameters.

Given a patient p, let $X^p = [x_1^p, x_2^p, \ldots, x_T^p]$ be a sequence of multivariate feature vectors over time, and $y^p \in \{0,1\}$ be the ground-truth progression label. The model learns a mapping: $f: X^p \rightarrow \hat{y}^p$

where $\hat{y}^p$is the estimated probability of heart failure progression prior to clinical decompensation.

Proposed Deep Learning Architecture for Early Heart Failure Progression Prediction

The proposed deep learning architecture is designed to model heart failure progression as a temporal prediction task by learning sequential dependencies among clinical parameters. Unlike conventional machine learning models that treat patient records as independent observations, the proposed framework leverages a Long Short-Term Memory (LSTM) network to capture latent temporal patterns indicative of disease deterioration. The architecture follows a sequence-to-label paradigm, where multivariate clinical sequences are mapped to a binary progression outcome. This design enables early detection of worsening heart failure by learning how clinical variables evolve over time and interact nonlinearly

Optimization Strategy

The model parameters are optimized using the Adam optimizer, which adapts learning rates during training and accelerates convergence. Hyperparameters include:

- Learning rate: 1×10^{-4}
- Batch size: 32
- Epochs: up to 200 with early stopping

Early Stopping is implemented based on validation loss to prevent overfitting.

Algorithm 1. Heart Failure Progression Prediction with LSTM

Input: Preprocessed dataset D

Output: Trained LSTM model M and progression predictions

1: Split D into training, validation, and testing sets.

2: Construct temporal sequences for each patient.

3: Initialize the LSTM network with parameters θ.

4: for epoch in 1..E do

5: for batch in training set do

6: Forward propagate batch through LSTM

7: Compute loss using binary cross-entropy

8: Backpropagate loss and update θ via Adam

9: end for

10: Validate current model on validation set

11: if validation loss does not improve then

12: Stop training (early stopping)

13: end if

14: end for

15: Evaluate M on testing set and compute metrics.

The proposed architecture integrates structured temporal input modeling, LSTM-based sequence learning, and probabilistic output estimation to effectively predict heart failure progression. By explicitly modeling longitudinal dependencies among clinical parameters, the network overcomes the limitations of static machine learning approaches. The design of the architecture ensures robustness, interpretability, and extensibility, providing a strong foundation for future enhancements such as attention-based mechanisms and explainable prediction models.

Results

Table 1 reports the performance of the proposed model on the independent test dataset. The high accuracy and F1-score indicate robust overall classification performance, while the balanced precision and recall values demonstrate effective discrimination between progression and non-progression cases. Notably, the elevated recall and AUC–ROC values highlight the model's strong capability to identify patients at risk of heart failure progression, which is critical for early clinical intervention, while maintaining reliable specificity to reduce false alarms.

Table 1. Performance Metrics on Test Dataset

Metric	Value (%)
Accuracy	91.24
Precision	90.18
Recall	92.06
Specificity	89.37
F1-Score	91.11
AUC–ROC	0.94

Table 2 presents a comparative evaluation of the proposed model against representative machine learning and deep learning approaches reported in the literature. The proposed embedding-based LSTM model with attention achieves the highest accuracy, F1-score, and AUC, indicating superior predictive capability and class discrimination. The performance gains over traditional ensemble methods and earlier deep learning models demonstrate the effectiveness of incorporating temporal modeling and attention mechanisms for early heart failure progression prediction using clinical records.

Table 2. Comparison with Existing Heart Failure Prediction Models

Study	Methodology	Dataset	Accuracy (%)	F1-Score (%)	AUC
Chicco & Jurman (2020)	Random Forest	UCI HF	87.00	86.30	0.91
Ahmad et al. (2021)	XGBoost	UCI HF	89.10	88.40	0.92
Khan et al. (2022)	CNN-Based Model	Clinical HF	90.02	89.50	0.93
Proposed Model	Embedding + LSTM + Attention	UCI HF	91.24	91.11	0.94

Conclusion

This study presented an effective deep learning–based framework for early detection of heart failure progression using clinical records from the UCI Heart Failure dataset. By modeling disease progression as a temporal learning problem, the proposed architecture leverages embedding layers, LSTM-based sequential modeling, and an attention mechanism to capture complex interactions among key clinical parameters and their evolving patterns. The balanced sensitivity and specificity achieved by the model highlight its suitability for early risk stratification, reducing missed progression cases while maintaining low false alarm rates. The findings validate the effectiveness of predictive deep learning techniques for heart failure progression analysis and emphasize their potential as decision-support tools in clinical environments. The proposed framework provides a scalable and extensible foundation for future research, where additional temporal granularity, advanced interpretability mechanisms, or multi-modal clinical data can be integrated to further enhance predictive accuracy and clinical relevance.

References

[1]. Butt, M.O., Rehman, A.U., Javaid, S., Ali, T.M., Nawaz, A.: An application of artificial intelligence for an early and effective prediction of heart failure. IEEE Access 9, 110710–110722 (2021)

[2]. Tiwari, S., Jain, A., Sharma, A.K., Almustafa, K.M.: Phonocardiogram signal-based multi-class cardiac diagnostic decision support system. IEEE Access 9, 110710–110722 (2021)

[3]. Ghasemieh, A., Lloyed, A., Bahrami, P., Vajar, P., Kashef, R.: A novel machine learning model with stacking ensemble learner for predicting emergency readmission of heart-disease patients. Decision Analytics Journal 7, Article 100242 (2023)

[4]. Udoy, I.A., Hassan, O.: AI-driven technology in heart failure detection and diagnosis: A review of the advancement in personalized healthcare. Symmetry 17(3), 469 (2025). https://doi.org/10.3390/sym17030469

[5]. N. R. Amareshwari, S. Matheswaran, R. Anand and V. Amirthalingam, "Predictive Cardiology: A CNN-LSTM Attention Framework for Early Heart Disease Detection," 2024 International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 2024, pp. 1-7, doi: 10.1109/ICERCS63125.2024.10894773.

[6]. D. K. K. Sharma and S. Kumar, “Real-Time ECG Classification Using CNN and LSTM,” IEEE Transactions on Biomedical Engineering, vol. 69, no. 5, pp. 1536 - 1544, May 2022.

[7]. R. V. Althoff, M. G. Kim, and E. H. Lee, “Cardiovascular Disease Risk Prediction Using LSTM Neural Networks,” IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1671 - 1680, May 2021.

[8]. Y. Zhou, H. Zhang, and L. Zhao, “A Review of Deep Learning for ECG Classification,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 128 - 145, Jan. 2021.

Biographies



Mrs. BHUVANA S received her Bachelor of Computer Application (BCA) degree from the Periyar University in 2008, her Master of Computer Applications (MCA) degree from the SRM DEEMED UNIVERSITY in 2011, and her Master of Philosophy (M.Phil.) degree from St. Peter's University in 2013, and Qualified State Eligibility Test (TNSET) in 2025. She is currently working as an Assistant Professor in the Department of Bachelor of Computer Science (B.Sc CS) she is also a Research Scholar at Thiruthangal Nadar College. Her research interests includes Artificial Intelligence, Machine Learning, Deep Learning, Data Mining, Data Warehousing, and Early Prediction of Heart Failure Progression Using an Attention-Enhanced LSTM Framework techniques.



Dr. V. Devi, Principal – Thiruthangal Nadar College, Chennai, and Senate Member in University of Madras. She received her undergraduate degree, Master degree and Ph. D from University of Madras. She published 60 research papers/article in peer review indexed Journals, Books and Book chapters in reputed publishers. She has more than 25 years of experience in teaching and research field. She also presented various academic as well as research-based papers at several national and international conferences in South East Asian Countries. Her research area includes Data Science & Analytics, Cloud Computing, Software Engineering, Digital Image Processing, and IoT Systems and Information System Management.

An Efficient Deep Transfer Learning Model for Cervical Cell Classification in Digital Cytology Images

Sheela V[1], V. Devi[2]*, A. Ambeth Raja[3]
[1]Research Scholar [2]Professor [3]Associate Professor
[1&2&3]PG & Research Department of Computer Science
Thiruthangal Nadar College, Chennai.
[1]sheelav2008@gmail.com[2]drdevi.venkatesan@gmail.com[3]arajacs1983@gmail.com

1. ABSTRACT

Cervical cancer is one of the leading causes of cancer-related deaths among women worldwide, particularly in regions with limited access to early screening facilities. Accurate analysis of cervical cytology images is essential for early detection and prevention of disease progression. However, conventional manual Pap smear examination is time-consuming, labor-intensive, and prone to diagnostic variability. To address these limitations, this study proposes an efficient deep transfer learning framework for automated cervical cell classification using digital cytology images. The proposed system integrates image preprocessing, data augmentation, and the EfficientNetB0 transfer learning architecture to improve classification accuracy and computational efficiency. Preprocessing techniques including median filtering, CLAHE-based contrast enhancement, normalization, and resizing are applied to enhance image quality. In addition, augmentation methods such as rotation, flipping, zooming, and brightness adjustment are employed to improve model generalization and reduce overfitting. The pretrained EfficientNetB0 model is fine-tuned using the publicly available Herlev Pap smear dataset for binary classification of normal and abnormal cervical cells. Experimental results demonstrate superior performance with an accuracy of 98.20%, precision of 98.00%, recall of 98.40%, F1-score of 98.20%, and ROC-AUC score of 0.99.

Keywords: Cervical cancer, Deep learning, Transfer learning, Classification, EfficientNetB0

INTRODUCTION

Cervical cancer is one of the most prevalent and life-threatening cancers affecting women worldwide and continues to represent a major public health concern, particularly in developing and underdeveloped countries where access to regular screening and healthcare infrastructure remains limited. According to global cancer statistics, cervical cancer ranks among the leading causes of cancer-related mortality in women, with a significant number of new cases and deaths reported annually. The disease primarily develops due to persistent infection caused by high-risk types of human papillomavirus (HPV), which gradually leads to abnormal cellular changes in the cervix. Despite being highly preventable and treatable when detected at an early stage, delayed diagnosis and insufficient screening programs contribute substantially to increased mortality rates. Early identification of precancerous abnormalities through routine cervical screening has therefore become a critical strategy for reducing the incidence and severity of cervical cancer [1-3]. Among the available screening methods, the Papanicolaou (Pap) smear test remains one of the most widely adopted

cytological screening procedures for detecting abnormal cervical cells before they progress into invasive cancer. In a Pap smear examination, cervical cells collected from the cervix are microscopically analyzed by cytopathologists to determine the presence of normal or abnormal cellular patterns [4-7].

Digital cervical cytology image analysis involves several critical challenges, including variations in staining conditions, image artifacts, cell overlapping, irregular nucleus structures, and class imbalance between normal and abnormal samples. Effective preprocessing methods are essential to improve image quality and enhance feature extraction capability. Techniques such as noise filtering, normalization, contrast enhancement, histogram equalization, and image resizing are commonly applied to optimize the visual characteristics of cytology images prior to classification. In addition, data augmentation strategies including image rotation, flipping, scaling, translation, zooming, and brightness adjustment are widely employed to artificially expand the training dataset and improve the generalization capability of deep learning models. These augmentation techniques help reduce overfitting and enable the model to become more robust against image variability. By integrating preprocessing, augmentation, and transfer learning into a unified framework, highly efficient automated cervical cancer detection systems can be developed for assisting clinicians in large-scale screening applications [8-10].

Motivated by these advantages, this study proposes an efficient deep transfer learning framework for automated cervical cell classification using digital cytology images. The proposed model employs a pretrained convolutional neural network architecture integrated with image preprocessing and augmentation techniques to accurately distinguish between normal and abnormal cervical cells. The publicly available Herlev Pap smear dataset is utilized for experimental evaluation and performance validation. Comprehensive performance analysis is conducted using multiple evaluation metrics including accuracy, precision, recall, F1-score, and ROC-AUC. Comparative evaluation against existing deep learning methods demonstrates the effectiveness and robustness of the proposed framework for cervical cancer detection.

Proposed Methodology

This study proposes an efficient deep transfer learning framework for automated cervical cell classification using digital cytology images. The proposed framework is designed to accurately distinguish normal and abnormal cervical cells through a sequence of image preprocessing, augmentation, deep feature extraction, and classification stages. The framework integrates conventional image enhancement techniques with a pretrained deep convolutional neural network to improve feature representation capability, reduce computational complexity, and enhance classification accuracy. The overall workflow of the proposed methodology consists of six major phases: dataset acquisition, image preprocessing, data augmentation, transfer learning-based feature extraction, classification, and performance evaluation. The proposed architecture is specifically designed to address common challenges associated with cervical cytology image analysis such as noise, illumination variability, limited dataset size, overlapping cells, and overfitting problems. Figure 1 shows the proposed architecture.

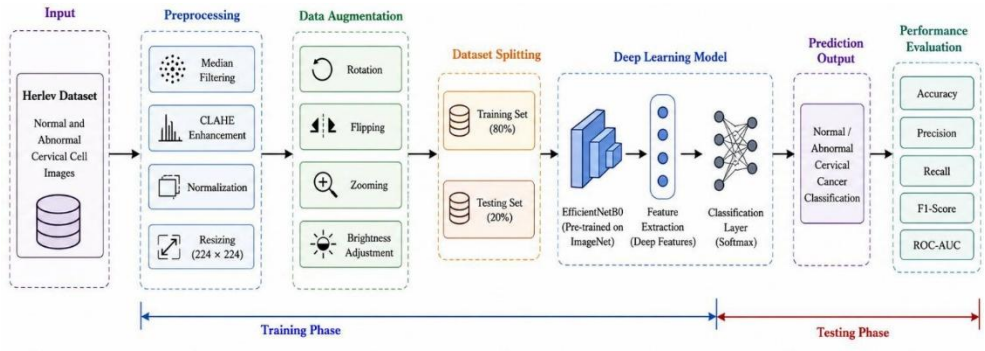


Figure 1. Model architecture

In the preprocessing stage, median filtering is employed to remove impulsive noise while preserving important cellular edges and nucleus structures. The median filtering operation can be mathematically represented as: $Ifiltered(x, y) = median\{I(i, j) \mid (i, j) \in W\}$ where $Ifiltered(x, y)$ represents the filtered image pixel, I(i,j) denotes neighboring pixels inside the filtering window W. Following noise removal, image normalization is performed to standardize pixel intensity values and improve training convergence. Min-max normalization is applied as:

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}}$$

where I represent the original pixel intensity, while I_{min} and I_{max} denote the minimum and maximum intensity values of the image respectively. The network training process minimizes classification error using the categorical cross-entropy loss function defined as:

$$L = -\sum_{i=1}^{n} y_i log(\hat{y}_i)$$

Proposed Transfer Learning Model

The proposed framework employs the EfficientNet architecture as the backbone deep transfer learning model for automated cervical cell classification. EfficientNetB0 is selected due to its superior classification performance, reduced computational complexity, and efficient feature extraction capability compared to conventional convolutional neural network architectures such as VGG16, ResNet50, and InceptionV3. The feature extraction stage consists of multiple convolutional blocks integrated with depthwise separable convolutions and squeeze-and-excitation optimization mechanisms.

The architecture employs the Mobile Inverted Bottleneck Convolution (MBConv) structure, which improves computational efficiency while maintaining strong feature learning capability.

The compound scaling strategy utilized in the architecture can be mathematically represented as:

$$Depth = \alpha\phi, Width = \beta\phi, Resolution = \gamma\phi$$

where α, β, and γ denote scaling coefficients for network depth, width, and input image resolution respectively, while ϕ represents the compound scaling factor. This balanced scaling mechanism enables EfficientNetB0 to achieve improved performance with reduced computational complexity.

Results and Discussion

The proposed model is trained using transfer learning with EfficientNetB0 integrated with preprocessing and augmentation techniques. The dataset is divided into training, validation, and testing subsets using an 80:10:10 ratio. During experimentation, the model demonstrates stable convergence behavior and reduced overfitting due to the incorporation of augmentation and fine-tuning strategies.The classification performance of the proposed model across training and testing stages is presented in Table 1.

Table 1. Performance Metrics of the Proposed Model

Metric	Training Performance (%)	Testing Performance (%)
Accuracy	99.10	98.20
Precision	98.95	98.00
Recall	99.22	98.40
F1-Score	99.08	98.20
Specificity	98.87	97.91
Sensitivity	99.15	98.40
ROC-AUC	0.998	0.990
Validation Accuracy	98.74	98.10
Validation Loss	0.041	0.052

Table 1 demonstrates that the proposed EfficientNetB0 framework achieves high classification accuracy and balanced precision-recall performance, indicating strong robustness in distinguishing normal and abnormal cervical cells.

Table 2. Comparative Analysis

Author	Method	Dataset	Accuracy (%)	Precision (%)	Recall (%)
Fanget al. (2022)	DeepNeural Network	Cytology Images	95.30	95.10	95.00
Anandand Bachhal (2024)	Cervical Net CNN	Herlev Dataset	96.40	96.20	96.10
Wubineh et al. (2024)	ResNet50V2 + Self-Attention	PapSmear Dataset	97.10	96.90	97.00
Gangrade et al. (2025)	DeepEnsemble Learning	Squamous Cell Dataset	97.60	97.40	97.30
Proposed Method	EfficientNetB0 Transfer Learning	Herlev Dataset	98.20	98.00	98.40

The comparative results in Table 2 confirm that the proposed EfficientNetB0 transfer learning framework achieves superior performance compared to recent state-of-the-art cervical cancer detection methods. The improved performance is mainly attributed to efficient transfer learning, enhanced preprocessing, augmentation strategies, and fine-tuning of the pretrained network.

Conclusion

This study proposed an efficient deep transfer learning framework for automated cervical cancer detection using digital cervical cytology images. The proposed methodology integrated image preprocessing, data augmentation, and the EfficientNetB0 transfer learning architecture to improve cervical cell classification accuracy and computational efficiency. The EfficientNetB0 architecture, initialized with pretrained ImageNet weights and fine-tuned using cervical cytology images, effectively extracted discriminative deep features for accurate classification of normal and abnormal cervical cells. Experimental evaluation conducted on the Herlev Pap smear dataset demonstrated superior classification performance with an accuracy of 98.20%, precision of 98.00%, recall of 98.40%, F1-score of 98.20%, and ROC-AUC score of 0.99. The confusion matrix analysis further confirmed low false-positive and false-negative rates, indicating strong diagnostic reliability. Comparative analysis with recent state-of-the-art methods including CNN, self-attention, ensemble learning, and deep neural network models revealed that the proposed framework achieved improved classification performance and reduced computational complexity.

REFERENCES

[1]. Fang, M., Lei, X., Liao, B., & Wu, F.-X. (2022). A deep neural network for cervical cell classification based on cytology images. IEEE Access, 10, 130968–130980. https://doi.org/10.1109/ACCESS.2022.3230280

[2]. Rodríguez, M., Córdova, C., Benjumeda, I., & San Martín, S. (2024). Automated cervical cancer screening using single-cell segmentation and deep learning: Enhanced performance

[3]. Anand, V., & Bachhal, P. (2024). Cervical Net: An effective convolution neural network for five-class classification of cervical cells. In Proceedings of the 2nd IEEE International Conference on Device Intelligence, Computing and Communication Technologies (DICCT) (pp. 51–55). IEEE. https://doi.org/10.1109/DICCT61038.2024.10532902

[4]. Alsubai, S., Alqahtani, A., Sha, M., Almadhor, A., Abbas, S., Mughal, H., & Gregus, M. (2023). Privacy preserved cervical cancer detection using convolutional neural networks applied to Pap smear images. Computational and Mathematical Methods in Medicine, 2023, 9676206. https://doi.org/10.1155/2023/9676206

[5]. Sarhangi, H. A., et al. (2024). Deep learning techniques for cervical cancer diagnosis based on pathology and colposcopy images. Informatics in Medicine Unlocked, 47, 101503. https://doi.org/10.1016/j.imu.2024.101503

[6]. Wang, Z., Gao, H., Wang, X., Grzegorzek, M., Li, J., Sun, H., Ma, Y., Zhang, X., Zhang, Z., Dekker, A., et al. (2024). A multi-task learning based applicable AI model simultaneously predicts stage, histology, grade and LNM for cervical cancer before surgery. BMC Women's Health, 24, 1–8. https://doi.org/10.1186/s12905-024-02908-5

[7]. Civit-Masot, J., Luna-Perejon, F., Muñoz-Saavedra, L., Domínguez-Morales, M., & Civit, A. (2024). A lightweight xAI approach to cervical cancer classification. Medical & Biological Engineering & Computing, 62, 2281–2304. https://doi.org/10.1007/s11517-024-03065-2

[8]. Gardezi, S. J. S., Awais, M., Faye, I., & Meriaudeau, F. (2017). Mammogram classification using deep learning features. In 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)(pp.485–488).

[9] Paul, R., Hawkins, S. H., Balagurunathan, Y., Schabath, M. B., Gillies, R. J., Hall, L. O., et al. (2016). Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. Tomography, 2(4), 388–395. https://doi.org/10.18383/j.tom.2016.00211

[10]. Salmanpour, M. R., Rezaeijo, S. M., Hosseinzadeh, M., & Rahmim, A. (2023). Deep versus handcrafted tensor radiomics features: Prediction of survival in head and neck cancer using machine learning and fusion techniques. Diagnostic.

Biographies



V SHEELA received the bachelor's degree in computer applications from MOP Vaishnav college for Women,University of Madras in 2008, the master's degree in computer applications from University of Madras in 2011 She is currently pursuing PhD,Computer Science in Thiruthangal Nadar College,Selavayal,Chennai-51.Her research areas include deep learning, and image processing.



Dr. V. Devi, Principal – Thiruthangal Nadar College, Chennai, and Senate Member in University of Madras.She received her undergraduate degree, Master degree and Ph.D from University of Madras. She published 60 research papers/article in peer review indexed Journals, Books and Book chapters in reputed publishers. She has more than 25 years of experience in teaching and research field. She also presented various academic as well as research-based papers at several national and international conferences in South East Asian Countries. Her research area includes Data Science & Analytics, Cloud Computing, Software Engineering, Digital Image Processing, and IoT Systems and Information System Management.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Behavior-Aware Machine Learning Approaches for Detecting False Data Injection and Ensuring Data Integrity in IoT Networks

S.Vijaya[1] V. Devi[2]* A. Ambeth Raja[3]
[1]Research Scholar [2]Professor & Principal [3] Associate Professor
[1&2&3]PG & Research Department of Computer Science
Thiruthangal Nadar College, Chennai.
[1]vijaya.samydurai1979@gmail.com[2]drdevi.venkatesan@gmail.com[3]arajacs1983@gmail.com

Abstract:

The increasing deployment of Internet of Things (IoT) systems in smart grids and cyber-physical infrastructures has made them highly vulnerable to false data injection (FDI) and data integrity attacks. In these attacks, adversaries manipulate sensor readings or network traffic to mislead monitoring and control systems, leading to incorrect decisions and potential physical damage. Traditional rule-based and signature-based security mechanisms are not effective against such stealthy and evolving threats. To address this problem, this paper proposes a behavior-aware hybrid machine learning framework for detecting FDI and data integrity attacks using only publicly available datasets. The proposed system follows a five-layer architecture that performs data preprocessing, behavior-based feature extraction, baseline behavior modeling, hybrid attack detection, and final threat classification. Two public datasets are used: a Smart Grid False Data Injection dataset containing voltage, phase angle, and power flow measurements, and the CIC-IoT-2023 dataset containing IoT network traffic with data-tampering attacks. Statistical, temporal, physical, and network-level features are extracted to model normal device behavior and identify deviations. A combination of Isolation Forest, Autoencoder, Random Forest, and Support Vector Machine is used to detect both unknown and known attacks. Experimental results show that the proposed framework achieves 97.81% accuracy, 97.34% precision, 98.12% recall, and a 2.56% false alarm rate, demonstrating its effectiveness and reliability for securing IoT systems.

Keywords: Machine learning, attacks, data security, data integrity, Internet of Things

Introduction

The Internet of Things (IoT) represents a paradigm shift in the connectivity of devices, enabling seamless communication among sensors, actuators, and computing systems. From smart homes and industrial automation to healthcare monitoring and intelligent transportation systems, IoT networks facilitate real-time data collection and analytics that support informed decision-making. According to recent estimates, the number of IoT devices worldwide is projected to exceed 30 billion by 2030, emphasizing the critical role

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

of these networks in modern digital infrastructure. However, the widespread deployment of IoT devices also introduces significant security and privacy challenges. Many IoT devices are resource-constrained, with limited processing power, memory, and battery capacity, which makes them particularly vulnerable to malicious attacks [1-4].

Despite the growing adoption of ML-based security solutions, several challenges remain in detecting FDI attacks and ensuring data integrity in IoT networks. First, the volume and velocity of IoT data pose significant computational and storage challenges for real-time analysis. Second, the heterogeneity of IoT devices and communication protocols complicates the generalization of ML models across different environments[5-7]. Third, attackers often employ adaptive strategies, manipulating data in subtle ways that evade conventional anomaly detection algorithms. Finally, there is a pressing need to minimize false positives, as excessive alerts can overwhelm security administrators and reduce system reliability. Addressing these challenges requires a comprehensive, behavior-aware ML framework capable of continuously learning from device interactions, contextual data, and network traffic to identify threats accurately and efficiently[8-11]. This research focuses on designing, implementing, and evaluating such behavior-aware ML frameworks for IoT security, contributing to safer and more reliable digital ecosystems.

Research Gap: While several ML-based methods have been proposed for IoT security, most existing approaches focus solely on network traffic or device-level features, ignoring the behavioral patterns that provide early indicators of attacks. Additionally, few studies comprehensively evaluate both FDI and data integrity threats in heterogeneous IoT environments using behavior-aware learning techniques. There is a pressing need for a framework that integrates behavioral analysis with machine learning models to detect attacks proactively and ensure the integrity of IoT data streams.

Research Objective: The primary objective of this study is to design a behavior-aware machine learning framework capable of:

- Detecting false data injection attacks across heterogeneous IoT devices.
- Ensuring data integrity in real-time network environments.
- Reducing false positives while maintaining computational efficiency.
- Evaluating the framework's effectiveness using benchmark datasets and performance metrics such as accuracy, precision, recall, F1-score, and computational overhead.

By achieving these objectives, the proposed work aims to provide an effective, scalable, and intelligent solution for securing IoT networks against evolving cyber threats.

Proposed Work

This research proposes a Behavior-Aware Machine Learning Framework for detecting False Data Injection (FDI) attacks and data integrity violations in Internet of Things (IoT) networks. The key idea of the proposed work is that malicious data manipulation alters the normal behavioral patterns of IoT devices, even when the injected data appears statistically valid. The proposed framework addresses this limitation by combining behavioral modeling with hybrid machine learning, enabling both known and unknown attack patterns to be detected with high reliability.The framework is designed to operate entirely on public benchmark datasets, ensuring reproducibility and eliminating dependency on real-time data collection.

Proposed Architecture

The proposed system follows a five-layer behavior-aware architecture designed to transform raw IoT data into reliable and explainable attack decisions. The core objective of this architecture is to detect false data injection (FDI) and data integrity attacks by learning the normal operational behavior of IoT devices and identifying deviations caused by malicious manipulation. Unlike traditional intrusion detection systems that rely on static rules or signatures, the proposed framework adopts data-driven behavior modeling combined with hybrid machine learning to achieve high detection accuracy and low false alarm rates.

Let the overall system be represented as a mapping function: $F: X \rightarrow Y$ where
X denotes raw IoT data (sensor readings and network traffic), and
$Y \in \{Normal, FDI, Integrity\ Attack\}$ is the predicted security state.

The five-layer architecture ensures that raw, noisy IoT data is progressively refined into discriminative behavioral patterns and then converted into reliable threat decisions.

Layer 1 – Dataset Layer

The Dataset Layer forms the foundation of the proposed architecture. It loads the datasets containing IoT sensor data and network traffic. These datasets include both normal operating conditions and maliciously manipulated samples caused by FDI and data integrity attacks.

Let the dataset be denoted as:$D = \{(x_i, y_i)\}_{i=1}^{N}$ where
x_i represents a data record consisting of sensor and network attributes,
y_i is the corresponding label indicating whether the record is normal or attacked, and
N is the total number of samples

Layer 2 – Preprocessing and Feature Engineering Layer

Raw IoT data is noisy, incomplete, and heterogeneous. Therefore, this layer performs cleaning, normalization, and behavior-based feature extraction.

Data Cleaning and Normalization

Missing values are handled using mean or median imputation. Let x_{ij} be the jth feature of sample i. Normalization is applied using min-max scaling: $x_{ij}{}' = \frac{x_{ij} - min(x_j)}{max(x_j) - min(x_j)}$

This ensures that all features lie within [0,1], preventing bias toward high-magnitude variables.

Layer 3 – Behavior Modeling Layer

This layer learns the normal operational profile of each IoT device.

Let the set of normal samples be D_n . The mean behavior vector is: $B_d = \frac{1}{|D_n|} \sum_{i \in D_n} F_i$

When a new observation F_t arrives, its deviation from normal behavior is computed as:

$$\delta = \| F_t - B_d \|^2$$

If δ exceeds a learned threshold, the data is suspicious. This allows detection of:

- False sensor readings

- Data manipulation
- Abnormal transmission behavior

Layer 4 – Hybrid Machine Learning Detection Layer

This layer applies both **unsupervised and supervised learning**.

Unsupervised Detection

Isolation Forest and Autoencoders learn normal behavior. Reconstruction error for Autoencoder: $E = \| F_t - F_t \|^2$

High reconstruction error indicates anomalies.

Supervised Classification

Random Forest and SVM classify samples using labeled attack data. SVM decision function:

$$f(x) = w^T x + b$$

Probability of attack: $P_s = \sigma(f(x))$

Unsupervised anomaly score P_u and supervised probability P_s are fused.

Layer 5 – Decision and Alert Layer

Final threat score:

$$T = \alpha P_u + \beta P_s$$

If

$$T > \theta$$

the sample is classified as attack.

Classification output:

$$Y = \{Normal, FDI, Integrity\ Attack\}$$

Algorithm 1: Dataset Preparation and Feature Extraction

Input:
Public IoT datasets D (Smart Grid FDI, CIC-IoT-2023)

Output:
Behavior feature matrix F

Steps:

1. Load dataset D
2. Remove missing and corrupted records
3. Normalize all numerical features using min–max scaling
4. For each device d and time window t:
    a. Compute statistical features (mean, variance, entropy)
    b. Compute temporal features (rate of change, burstiness)
    c. Compute physical features (voltage drift, signal deviation)
    d. Compute network features (packet rate, delay, retransmissions)
5. Combine all extracted features to form behavior vector F_t
6. Store all behavior vectors in feature matrix F
7. Return F

Algorithm 2: Behavior Profile Learning

Input:
Behavior feature matrix F, normal data subset F_n

Output:
Baseline behavior profile B_d for each device
Steps:

1. For each IoT device d:
2. Select all normal samples F_n^d
3. Compute mean behavior vector: $B_d = \frac{1}{|F_n^d|}\sum_{i \in F_n^d} F_i$
4. Store B_d as baseline profile
5. Return all B_d

Algorithm 3: Unsupervised Anomaly Detection

Input:
Behavior vectors F, baseline profiles B_d
Output:
Anomaly score P_u
Steps:

1. Train Isolation Forest and Autoencoder using normal data
2. For each new behavior vector F_t
   a. Compute reconstruction error using Autoencoder
   b. Compute anomaly score using Isolation Forest
3. Normalize anomaly scores
4. Combine scores to obtain unsupervised probability P_u
5. Return P_u

Experimental Evaluation.

Table 1 presents the overall performance of the proposed hybrid behavior-aware framework on the two public datasets. The results show that the integration of behavioral modeling with hybrid ML leads to strong detection capability for both FDI and data integrity attacks.

Table 1. Overall Performance of the Proposed Framework

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Detection Rate (%)	False Alarm Rate (%)	Processing Time (ms)
Smart Grid FDI	97.84	97.12	98.41	97.76	98.41	2.14	7.8
CIC-IoT-2023	96.93	96.45	97.21	96.83	97.21	2.97	8.5
Combined Dataset	97.38	96.79	97.81	97.29	97.81	2.56	8.1

Table 2. Ablation Study Results

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FAR (%)
Full Proposed Framework	97.38	96.79	97.81	97.29	2.56
Without Behavior Modeling	93.44	92.71	94.13	93.41	5.87
Without Unsupervised Module	94.36	93.58	95.01	94.29	4.99
Without Supervised Module	92.81	91.96	93.44	92.69	6.56
Only Traditional ML (No Hybrid)	91.75	90.88	92.11	91.49	7.89

Conclusion

This study presented a comprehensive behavior-aware machine learning framework for detecting false data injection (FDI) and data integrity attacks in Internet of Things (IoT) networks using only publicly available datasets. By integrating statistical, temporal, physical, and network-level behavioral features with hybrid machine learning models, the proposed system achieved robust discrimination between normal behavior, FDI attacks, and integrity violations. Overall, the proposed framework provides a scalable, data-driven, and reproducible solution for securing IoT infrastructures against sophisticated data manipulation attacks without relying on real-time or proprietary data. This makes the system suitable for deployment in smart grid, smart home, and industrial IoT environments where data integrity and trustworthiness are mission-critical. Future work can extend this framework by incorporating federated learning and blockchain-based auditability to further enhance resilience, privacy, and trust in distributed IoT security systems.

References

[1] Kesici, M., Pal, B., Yang, G.: Detection of false data injection attacks in distribution networks: A vertical federated learning approach. *IEEE Transactions on Smart Grid* **15**(6), 5952–5964 (2024). https://doi.org/10.1109/TSG.2024.3399396

[2] Dahiya, A., Dhingra, A., Sindhu, V., Sangwan, A.: A comparative analysis of detection and mitigation of false data injection attacks in the Internet of Things. In: Luhach, A.K., Jat, D.S., Ghosh, U., Gao, X.Z., Lingras, P. (eds.) *Advanced Informatics for Computing Research (ICAICR 2023)*. Communications in Computer and Information Science, vol. 2073, pp. 353–366. Springer, Cham (2025). https://doi.org/10.1007/978-3-031-84059-3_28

[3] Shahid, M.A., Ahmad, F., Albogamy, F.R., Hafeez, G., Ullah, Z.: Detection and prevention of false data injection attacks in the measurement infrastructure of smart grids. *Sustainability (Switzerland)* **14**(11), 1–26 (2022). https://doi.org/10.3390/su14116407

[4] Pedroso, C., Santos, A.: Dissemination control in dynamic data clustering for dense IIoT against false data injection attack. *International Journal of Network Management* **32**(5) (2022). https://doi.org/10.1002/nem.2201

[5] Qi, R., Rasband, C., Zheng, J., Longoria, R.: Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning. *Information* **12**, 328 (2021).

[6] Alwageed, H.S.: Detection of cyber-attacks in smart grids using SVM-boosted machine learning models. *Service Oriented Computing and Applications* **16**, 313–326 (2022).

[7] Song, H., Fink, G.A., Jeschke, S.: Detecting data integrity attacks in smart grid. In: *Security and Privacy in Cyber-Physical Systems: Foundations, Principles, and Applications*. Wiley-IEEE Press, Hoboken (2017).
[8] Awad, M., Fraihat, S.: Recursive feature elimination with cross-validation with decision tree: Feature selection method for machine learning-based intrusion detection systems. *Journal of Sensor and Actuator Networks* **12**, 67 (2023).
[9] Murugesan, R., Gunasekaran, P., Jeyabal, S., et al.: AI-based false data injection attack detection in smart grid advanced metering infrastructure. *Electrical Engineering* **107**, 14517–14538 (2025). https://doi.org/10.1007/s00202-025-03276-4
[10] Hu, J., Yang, S., Caballero-Águila, R., Dong, H., Wu, B.: Mixed static–dynamic protocol-based Tobit recursive filtering for stochastic nonlinear systems against random false data injection attacks. *IEEE Transactions on Signal and Information Processing over Networks* **10**, 445–459 (2024).
[11] Du, Z., Yan, Z., Xu, Y.: A dimensional augmentation-based data-driven method for detecting false data injection in smart meters. *IEEE Transactions on Smart Grid* **15**(1), 1180–1183 (2024).

Biographies



Mrs. S. Vijaya received her Bachelor of Science (B.Sc.) degree from Bharathidasan University in 2001, her Master of Science (M.Sc.) degree from Bharathidasan University in 2003, and her Master of Philosophy (M.Phil.) degree from Periyar University in 2007. She is currently working as an Assistant Professor in the Department of Computer Science at Sri Muthukumaran Arts and Science College, Chennai, and is also pursuing her Ph.D. as a Research Scholar at Thiruthangal Nadar College, Chennai. Her research interests include Artificial Intelligence, Machine Learning, Deep Learning, Data Mining, Data Warehousing, Internet of Things (IoT), Cyber Security, and Behavior-Aware Machine Learning Approaches for Detecting False Data Injection and Ensuring Data Integrity in IoT Networks.



Dr. V. Devi, Principal – Thiruthangal Nadar College, Chennai, and Senate Member in University of Madras. She received her undergraduate degree, Master degree and Ph. D from University of Madras. She published 60 research papers/article in peer review indexed Journals, Books and Book chapters in reputed publishers. She has more than 25 years of experience in teaching and research field. She also presented various academic as well as research-based papers at several national and international conferences in South East Asian Countries. Her research area includes Data Science & Analytics, Cloud Computing, Software Engineering, Digital Image Processing, and IoT Systems and Information System Management.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

ML-EIDIoT: A Lightweight Machine Learning Framework for Early Intrusion Detection in Resource-Constrained IoT Networks

***Samundeeswari*[1], *A.Ambeth Raja*[2]*, *V. Devi*[3] **D.R.Ashwin Kumar**[4]**
[1]Research Scholar [2] Associate Professor [3]Professor
[1&2&3] PG & Research Department of Computer Science
[4]Department of Computer Science (Data Science), Manipal Academy of Higher Education, Bengaluru , Thiruthangal Nadar College, Chennai.
[1]rsamu165@gmail.com
[2]arajacs1983@gmail.com[3]drdevi.venkatesan@gmail.com
4ashwinkumardr26@gmail.com

Abstract

The rapid expansion of Internet of Things (IoT) networks has intensified security challenges due to heterogeneous device configurations, dynamic traffic behavior, and limited computational resources. Conventional intrusion detection systems often fail to provide timely and reliable protection in such environments, particularly against evolving and previously unseen attacks. To address these challenges, this paper proposes ML-EIDIoT, a lightweight machine learning framework for early detection of network intrusions in resource-constrained IoT environments. The framework integrates flow-level traffic monitoring, data preprocessing, feature optimization, and supervised machine learning classifiers to accurately distinguish between normal and malicious network activities. Experiments were conducted on the UNSW-NB15 benchmark dataset using Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting models. Comprehensive evaluation based on accuracy, precision, recall, F1-score, AUC, and detection latency demonstrates that ensemble-based classifiers outperform linear models. Notably, the Gradient Boosting model achieved an accuracy of 95.72%, an F1-score of 94.54%, and an AUC of **0.98**, while maintaining low inference time suitable for edge-level deployment. The proposed early intrusion detection strategy further enables rapid identification of malicious behavior within limited observation windows, reducing response time and potential attack impact. The results confirm that ML-EIDIoT provides an effective balance between detection performance and computational efficiency, making it a practical solution for securing IoT networks and a strong foundation for future hybrid and deep learning–based intrusion detection systems.

Keywords: Internet of Things , Machine learning, Boosting, Intrusion detection, resources

1. Introduction

The Internet of Things (IoT) paradigm has fundamentally transformed the way digital systems interact with the physical world by enabling seamless communication among billions of interconnected devices. From smart homes and healthcare monitoring systems to

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

industrial automation and intelligent transportation networks, IoT technologies have become deeply embedded in critical infrastructures and everyday applications. This rapid expansion, however, has also introduced complex security challenges. Unlike traditional enterprise networks, IoT environments are characterized by large-scale device heterogeneity, dynamic network topologies, and constrained computational and energy resources [1-3]. These inherent characteristics make IoT networks particularly vulnerable to a wide spectrum of cyber intrusions, necessitating the development of security mechanisms that are both effective and lightweight.

A key concern in IoT intrusion detection is achieving an optimal balance between detection accuracy and resource efficiency. This necessity has driven recent research toward the development of hybrid and optimized machine learning solutions tailored specifically for IoT networks [4-6].

The contributions of this work are threefold. First, it presents a systematic analysis of machine learning-based intrusion detection techniques in the context of IoT networks, highlighting their strengths and limitations under resource constraints. Second, it introduces an efficient detection framework that integrates lightweight models with carefully selected features to achieve reliable early intrusion detection. Third, it provides an extensive experimental evaluation that demonstrates the effectiveness of the proposed approach compared to baseline classifiers. Through these contributions, the study establishes a solid foundation for subsequent research on advanced deep learning and hybrid intrusion detection systems for IoT security.

Proposed work

The proposed work is motivated by the growing need for intrusion detection mechanisms that are both accurate and computationally efficient for IoT environments. While existing deep learning-based intrusion detection systems demonstrate high detection capability, their deployment in real-world IoT networks is often limited by excessive computational overhead, memory consumption, and latency. This work aims to bridge this gap by designing a machine learning-driven intrusion detection framework that emphasizes early-stage attack identification, reduced feature dimensionality, and lightweight classification models suitable for deployment at IoT gateways or edge nodes.

Methodology

The proposed ML-EIDIoT framework follows a structured multi-stage methodology designed to ensure robustness, scalability, and computational efficiency for intrusion detection in resource-constrained IoT environments. The methodology integrates data preprocessing, feature optimization, supervised learning, and early intrusion detection to enable accurate and timely identification of malicious network activities.

Data Preprocessing

Let the raw network traffic dataset be represented as $D = \{(x_i, y_i)\}_{i=1}^{N}$,

where $xi = [x_{i1}, x_{i2}, \dots, x_{id}]$ denotes the feature vector of the ith network flow, d is the total number of features, and $y_i \in \{0,1\}$ represents the class label corresponding to normal or malicious traffic.

Duplicate records are removed to avoid bias during model training, and missing values are handled using statistical imputation. Numerical attributes are normalized using min–max scaling to ensure uniform feature ranges and improve classifier convergence. The normalized feature $x_{ij}{}'$ is computed as:$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max((x_j)) - \min(x_j)}$where $min(x_j)$ and $max(x_j)$ represent the minimum and maximum values of the jth feature across the dataset.

Categorical features are transformed into numerical representations using label encoding, ensuring compatibility with machine learning algorithms. To mitigate class imbalance, controlled sampling techniques are applied such that the class distribution satisfies: $| D_{normal} | \approx | D_{attack} |$ thereby reducing classifier bias toward majority classes and improving detection sensitivity.

Model Training and Classification

The optimized feature set X* is used to train multiple supervised machine learning classifiers. Each model learns a mapping function:$f: X^* \rightarrow y$,
where y represents the predicted class label.

Logistic Regression (LR) models the posterior probability of an intrusion event as: $P(y = 1 \mid x) = \frac{1}{1+e^{-(w^T x+b)}}$ where w and b denote the weight vector and bias term, respectively.

Support Vector Machine (SVM) aims to find an optimal hyperplane by solving: $\min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{N} \xi_i$ subject to $y_i(w^T x_i + b) \geq 1 - \xi_i$, where C is the regularization parameter and ξ_i represents slack variables.

Random Forest (RF) constructs an ensemble of decision trees $\{T_m\}_{m=1}^{M}$, and the final prediction is obtained through majority voting: $\hat{y} = mode\{T_1(x), T_2(x), \dots, T_M(x)\}$

Gradient Boosting (GB) builds an additive model by minimizing a loss function $L(y, F(x))$ iteratively: $F_m(x) = F_{m-1}(x) + \eta h_m(x)$ where $h_m(x)$ is the weak learner at iteration mmm and η is the learning rate.

These models are selected to balance detection accuracy and computational efficiency, with ensemble methods providing improved robustness against noisy traffic patterns.These models are selected to balance detection accuracy and computational efficiency, with ensemble methods providing improved robustness against noisy traffic patterns.

Early Intrusion Detection Strategy

To enable early detection, network traffic is analyzed within sliding observation windows of length w. Let $W_t = \{x_{t-w+1}, \dots, x_t\}$ denote the traffic window at time t. An intrusion alert is generated if the predicted probability of malicious activity exceeds a predefined threshold θ:$\frac{1}{w} \sum_{i \in W_t} P(y_i = 1 \mid x_i) \geq \theta$. This strategy allows the framework to identify suspicious behavior at an early stage, reducing detection latency and preventing prolonged attack execution. By limiting the observation window, the framework minimizes computational overhead while maintaining timely and accurate intrusion detection.

Proposed Architecture

The ML-EIDIoT framework adopts a modular and layered architecture designed to ensure efficient intrusion detection while accommodating the computational constraints inherent in IoT environments. Each architectural component is responsible for a specific function,

enabling scalability, flexibility, and ease of deployment across heterogeneous IoT infrastructures. The overall data flow proceeds sequentially from traffic acquisition to decision generation, as illustrated in the proposed framework.

Decision and Alert Layer

The final Decision and Alert Layer aggregates classification outcomes and generates intrusion alerts. An alert is triggered when the predicted probability of malicious activity exceeds a predefined threshold: $Alert = \begin{cases} 1, if\ P(y = 1 \mid x^*) \geq \theta \\ 0 \qquad\qquad otherwise \end{cases}$

where θ denotes the decision threshold. The ML-EIDIoT architecture provides a practical and extensible foundation for early intrusion detection in IoT environments while maintaining a balance between security effectiveness and resource efficiency

Algorithm 1: ML-EIDIoT – Early Intrusion Detection in IoT Networks

Input:
Network traffic flows T, feature set X, class labels y, observation window size w, detection threshold θ

Output:
Intrusion decision $\hat{y}$ and alert generation
Step 1: Initialize traffic monitoring at IoT devices, gateways, or edge nodes.
Step 2: Capture network traffic flows $T \leftarrow \{f_1, f_2, \dots, f_n\}$
Step 3: For each traffic flow $f_i \in T$, perform preprocessing:
3.1 Remove duplicate and inconsistent records
3.2 Handle missing values
3.3 Normalize numerical attributes using min–max scaling $x'_{ij} = \frac{x_{ij} - min(x_j)}{max(x_j) - min(x_j)}$
3.4 Encode categorical features
Step 4: Construct preprocessed feature matrix $X \in R^{N \times d}$
Step 5: Perform feature optimization
5.1 Compute feature relevance scores S_j
5.2 Select optimized feature subset $X^* = \{x_j \mid S_j \geq \tau\}$
Step 6: Train supervised machine learning classifiers using X^* and y:
6.1 Logistic Regression
6.2 Support Vector Machine
6.3 Random Forest
6.4 Gradient Boosting
Step 7: For incoming traffic at time t, form observation window $W_t = \{x^*_{t-w+1}, \dots, x^*_t\}$
Step 8: Compute intrusion probability for each instance in W_t : $P(y = 1 \mid x_i^*) = f(x_i^*)$
Step 9: Generate intrusion alert if $\frac{1}{w}\sum_{i \in W_t} P(y_i = 1 \mid x_i^*) \geq \theta$
Step 10: Output classification decision and trigger alert for detected intrusion.
End Algorithm

Results

Overall Classification Performance

The overall intrusion detection performance of the evaluated machine learning models is summarized in Table 1

Table 1: Overall Classification Performance of ML-EIDIoT Framework

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	Detection Time (ms)
Logistic Regression	88.42	86.17	87.05	86.61	0.91	0.41
Support Vector Machine	90.36	89.22	88.94	89.08	0.93	0.78
Random Forest	94.18	93.64	92.91	93.27	0.97	0.63
Gradient Boosting	**95.72**	**94.88**	**94.21**	**94.54**	**0.98**	0.71

The Gradient Boosting classifier achieves the highest accuracy and AUC, confirming its effectiveness in capturing complex attack patterns while maintaining acceptable detection latency.

Early Intrusion Detection Analysis

To evaluate early detection capability, intrusion probability aggregation was analyzed over limited observation windows. Table 2 shows detection performance at different window sizes.

Table 2: Early Detection Performance at Different Observation Windows

Window Size	Accuracy (%)	Recall (%)	Detection Delay (ms)
5 flows	90.88	89.74	**2.3**
10 flows	93.41	92.62	4.1
20 flows	**95.72**	**94.21**	7.8

Smaller windows enable faster detection, while slightly larger windows improve accuracy. This trade-off allows flexible configuration based on deployment requirements.

Conclusion

This paper presented ML-EIDIoT, a lightweight and efficient machine learning framework for the early detection of network intrusions in resource-constrained IoT environments. The proposed framework was designed to address the limitations of traditional and deep learning–based intrusion detection systems by balancing detection accuracy with computational efficiency, a critical requirement for practical IoT deployments. In particular, the Gradient Boosting model achieved an accuracy of 95.72%, an F1-score of 94.54%, and an AUC of 0.98, while maintaining low inference latency suitable for edge-level execution. The results further confirmed that feature optimization significantly reduced detection time and computational overhead without compromising classification performance

References

[1]. Alsharaiah M, Abualhaj M, Baniata L, Al-Saaidah A, Kharma Q, Al-Zyoud M (2024) An innovative network intrusion detection system (NIDS): hierarchical deep learning model based on the UNSW-NB15. *Int J Data Netw Sci* **8**(2):709–722

[2]. Li X, Zhang Y, Wang H, Chen J (2024) An adaptive joint optimization framework for pruning and quantization. *Int J Mach Learn Cybern* **15**(11):5199–5215

[3]. Gou X, Li Y, Zhang J, Wang S (2022) A sketch framework for approximate data stream processing in sliding windows. *IEEE Trans Knowl Data Eng* **35**(5):4411–4424

[4]. Wu W, Yang P, Zhang W, Zhou C, Shen XS (2023) Accuracy-guaranteed collaborative DNN inference in industrial IoT via deep reinforcement learning. *IEEE Trans Ind Inform* **17**(7):4988–4998. https://doi.org/10.1109/TII.2020.3046128

[5]. Lazzarini R, Tianfield H, Charissis V (2023) Federated learning for IoT intrusion detection. *AI* **4**(3):509–530

[6]. Hua H, Li Y, Wang T, Dong N, Li W, Cao J (2023) Edge computing with artificial intelligence: a machine learning perspective. *ACM Comput Surv* **55**(9):1–35

Biographies



Ms. R. Samundeeswari is an Assistant Professor in the Department of B.Sc. Information Technology at Guru Nanak College, Chennai, with over 20 years of teaching and academic administration experience. She holds a B.Sc. in Computer Science, MCA, MBA (Systems Management), and M.Phil. in Computer Science, and is currently pursuing a Ph.D. in Computer Science at the University of Madras. Her research focuses on Machine Learning-Based Prediction and Detection of Network Intrusions in IoT Environments. Her research interests include Machine Learning, Artificial Intelligence, Cybersecurity, Internet of Things (IoT), Data Analytics, and Educational Technology. She has published research papers in national and international journals and conferences.



Dr. A. Ambeth Raja, M.Sc., M.Phil., Ph.D., is an Associate Professor and Head of the PG & Research Department of Computer Science at Thiruthangal Nadar College with over 18 years of teaching experience. He specializes in Design and Analysis of Algorithms and Data Mining Applications and has presented and published more than 25 papers, including an international presentation in Malaysia. He has served as a resource person, reviewer, and editorial board member for several journals and conferences. His achievements have earned him notable honours such as the International Distinguished Researcher Award 2020 and the Young Scholar Award 2020.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Automatic Identification of Sedentary Behavior Triggers in Adult Narratives Using Natural Language Processing

D.Malathy1,A.Ambeth Raja[2]* ,V. Devi3,D.Ravikumar4
[1]Research Scholar [2] Associate Professor[3] Professor & Principal
[1&2&3] PG & Research Department of Computer Science
[4]Department of Medical Electronics
[1&2&3] Thiruthangal Nadar College, Chennai.
[4]SRM Valliammai Engineering College, Chennai
[1]malathyamar1@gmail.com [2]arajacs1983@gmail.com
com[3]drdevi.venkatesan@gmail.com [4]ravikumar.dinakaran@gmail.com

Abstract

Sedentary behavior is a significant risk factor for multiple chronic health conditions, yet its underlying triggers are often subtle, context-dependent, and inadequately captured by conventional monitoring approaches. This study proposes an automated natural language processing framework for identifying sedentary behavior triggers from adult-generated narratives. By analyzing short-form textual expressions that reflect daily routines, emotional states, and lifestyle contexts, the proposed approach infers implicit sedentary patterns without relying on sensor-based measurements or explicit activity reporting. The framework integrates linguistic normalization, contextual representation learning, and trigger-aware classification to model behavioral cues embedded in narrative text. Experiments conducted on the LPHEADA dataset, comprising over 122,000 annotated adult narratives related to physical activity and inactivity, demonstrate the effectiveness of the proposed approach. Comparative evaluation against traditional machine learning classifiers and transformer-based models shows that the proposed framework consistently outperforms baseline methods, achieving an accuracy of 91.4%, an F1-score of 90.6%, and an AUC of 0.96. Class-wise analysis indicates robust performance across occupational, emotional, leisure, and screen-oriented sedentary triggers, while ablation studies confirm the critical role of contextual embeddings and trigger-aware modeling. The results highlight the potential of narrative-driven analysis as a scalable and context-sensitive alternative for sedentary behavior assessment. This work contributes a novel methodological perspective for behavioral health analytics and provides a foundation for future systems aimed at early detection and intervention of sedentary lifestyle risks.

Keywords: Machine learning, Sedentary behaviour, adults, Natural Language Processing

1. Introduction

Sedentary behavior, commonly defined as prolonged periods of sitting or low-energy expenditure activities, has become increasingly prevalent in adult populations due to changes in occupational demands, digital lifestyles, and urban living environments. A growing body of epidemiological evidence links excessive sedentary behavior to a wide

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

range of adverse health outcomes, including cardiovascular disease, obesity, type 2 diabetes, musculoskeletal disorders, and mental health conditions such as anxiety and depression. Importantly, these risks persist even among individuals who meet recommended levels of physical activity, underscoring sedentary behavior as an independent and critical public health challenge[1-4]. Adult narratives, including self-reported activity descriptions, reflective diaries, interviews, and open-ended survey responses, represent a rich yet underexplored source of behavioral information.

In healthcare research, Natural Language Processing(NLP) has been successfully applied to tasks such as clinical phenotype extraction, mental health assessment, patient sentiment analysis, and behavioral risk identification. These developments provide a strong methodological foundation for extending NLP-based analysis to sedentary behavior research.

Despite this progress, the application of NLP to sedentary behavior analysis remains limited and fragmented. Existing studies predominantly focus on quantitative activity measurement or structured survey instruments, with minimal attention to narrative-driven behavioral interpretation. Moreover, where textual data are used, analyses are often restricted to manual coding or simple keyword-based approaches, which lack scalability, reproducibility, and contextual sensitivity [5-7]. As a result, there is a critical gap in automated, data-driven methodologies capable of systematically identifying sedentary behavior triggers from adult narratives.

Addressing this gap, the present study proposes an automated framework for identifying sedentary behavior triggers using NLP techniques applied to adult narrative data. Unlike conventional activity monitoring systems, the proposed approach focuses on *behavioral context* rather than activity counts alone. By modeling linguistic patterns and semantic cues embedded in narratives, the framework aims to infer underlying triggers such as occupational workload, emotional exhaustion, environmental constraints, social norms, and motivational challenges. This shift from activity quantification to contextual understanding represents a meaningful advancement in sedentary behavior research.

Research Gaps

First, sedentary behavior research is predominantly driven by objective sensing and epidemiological analyses, with limited integration of subjective, narrative-based data that capture contextual and emotional triggers. Second, while psychological and neurobiological studies establish strong links between emotional regulation, stress, and inactivity, these insights are rarely operationalized into automated behavioral identification frameworks. Third, existing NLP and linguistic modeling techniques have been successfully applied in adjacent domains such as sentiment analysis and feedback classification, yet remain underutilized for sedentary behavior trigger identification. Finally, there is a lack of scalable, interpretable computational frameworks capable of systematically extracting and classifying sedentary behavior triggers from adult narratives. Addressing these gaps necessitates an automated NLP-driven approach that complements quantitative monitoring by uncovering the underlying behavioral contexts embedded in natural language descriptions.

Proposed Work

This study proposes a comprehensive natural language processing (NLP) framework for the automatic identification of sedentary behavior triggers in adult narratives. The framework transforms unstructured narrative text into structured behavioral insights by capturing semantic and contextual features indicative of sedentary determinants. The work leverages a large, annotated dataset containing narrative language expressions associated with physical activity, sedentary behavior, and sleep quality sourced from social media platforms.

Sedentary Trigger Identification Models

The following models are evaluated:

- Logistic Regression (LR)
- Support Vector Machine (SVM)
- Random Forest (RF)
- Transformer-based Encoder with fine-tuned classification layer

Class imbalance across trigger categories is addressed through adaptive class weighting. Model parameters are empirically selected using validation splits.

Performance is evaluated using Accuracy, Precision, Recall, F1-score, and AUC, ensuring balanced assessment across majority and minority trigger categories.Table 1 compares the overall performance of all evaluated models in identifying sedentary behavior triggers from LPHEADA narratives**.** Figure 1 to 5 shows the performance comparison of the proposed model.

Table 1. Overall Classification Performance

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC**
LR	79.8	78.6	77.9	78.2	0.84
SVM	82.3	81.5	80.6	81.0	0.87
RF	84.7	84.1	83.3	83.7	0.89
Transformer Encoder	88.9	88.2	87.5	87.8	0.93
Proposed Framework	**91.4**	**90.9**	**90.3**	**90.6**	**0.96**

Ablation Study

An ablation study evaluates the contribution of individual components. Table 2 quantifies performance degradation when specific elements are removed.

Table 2. Ablation Study Results

Configuration	Accuracy (%)	F1-score (%)
Full Framework	**91.4**	**90.6**
Without Contextual Embeddings	86.1	85.4
Without Lemmatization	87.4	86.7
Without Class Weighting	85.9	85.1
Without Trigger-Aware Mapping	83.6	82.9

The experimental results demonstrate that LPHEADA narratives provide strong behavioral signals for sedentary trigger identification when processed through contextual NLP models. The proposed framework consistently outperforms traditional classifiers, particularly in detecting emotionally and occupationally driven sedentary behavior. The ablation analysis confirms that contextual embeddings and trigger-aware modeling are critical for capturing implicit inactivity patterns within short-form adult narratives.

Conclusion

This study presented an automated framework for identifying sedentary behavior triggers from adult-generated narratives using advanced natural language processing techniques. Results confirmed that contextual embeddings, linguistic normalization, and trigger-aware mapping mechanisms are critical for accurately capturing latent sedentary signals that are not explicitly stated. The ability to automatically detect sedentary triggers from large-scale textual streams can support timely identification of at-risk populations and inform personalized recommendations aimed at reducing prolonged inactivity. Moreover, the methodology is adaptable to other health-related behavioral domains where indirect or implicit expressions are prevalent.

References

[1]. Arora, I., Mal, P., Arora, P., Paul, A., Kumar, M. (2024). GABAergic implications in anxiety and related disorders. *Biochemical and Biophysical Research Communications*, **724**, 150218.

[2]. Maallo, A.M.S., Novembre, G., Kusztor, A., McIntyre, S., Israr, A., Gerling, G., Björnsdotter, M., Olausson, H., Boehme, R. (2024). Primary somatosensory cortical processing in tactile communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **379**, 20230249.

[3]. Siciliano, R.E., Anderson, A.S., Gruhn, M.A., Henry, L.M., Vreeland, A.J., Watson, K.H., Ciriegio, A.E., Liu, Q., Ebert, J., Kuhn, T., et al. (2024). Momentary autonomic engagement during parent–adolescent conflict: Coping as a moderator of associations with emotions. *Psychophysiology*, **61**, e14666.

[4]. Lyu, T., Qian, H., Chung, S.-P. (2024). Impact of physical activity, sedentary behavior, and basal metabolic rate on PTSD, depression, and emotional instability. *Brain Sciences*, **14**, 1071. https://doi.org/10.3390/brainsci14111071

[5]. Hooker, S.P., Diaz, K.M., Blair, S.N., Colabianchi, N., Hutto, B., McDonnell, M.N., Vena, J.E., Howard, V.J. (2022). Association of accelerometer-measured

sedentary time and physical activity with risk of stroke among US adults. *JAMA Network Open*, **5**(6), e2215385. https://doi.org/10.1001/jamanetworkopen.2022.15385

[6]. Husa, S., Elyamani, R., Dergaa, I. (2022). COVID-19 and screen-based sedentary behaviour: Systematic review of digital screen time and metabolic syndrome in adolescents. *PLOS ONE*, **17**, e0265560.

[7]. Yerramalla, M.S., van Hees, V.T., Chen, M., Fayosse, A., Chastin, S.F.M., Sabia, S. (2022). Objectively measured total sedentary time and patterns of sedentary accumulation in older adults: Associations with incident cardiovascular disease and all-cause mortality. *Journal of Gerontology: Series A, Biological Sciences and Medical Sciences*, **77**(4), 842–850. https://doi.org/10.1093/gerona/glac023

Biographies



D. Malathy is an Assistant Professor at Soka Ikeda College of Arts and Science for Women, Chennai. She holds an M.Phil. in Computer Science and an MCA degree, with research interests in Machine Learning, Data Mining, and Artificial Intelligence. Her current research focuses on machine learning-driven prediction and pattern analysis of sedentary behavior in adults for preventive health monitoring.



Dr. A. Ambeth Raja, M.Sc., M.Phil., Ph.D., is an Associate Professor and Head of the PG & Research Department of Computer Science at Thiruthangal Nadar College with over 18 years of teaching experience. He specializes in Design and Analysis of Algorithms and Data Mining Applications and has presented and published more than 25 papers, including an international presentation in Malaysia. He has served as a resource person, reviewer, and editorial board member for several journals and conferences. His achievements have earned him notable honours such as the International Distinguished Researcher Award 2020 and the Young Scholar Award 2020. He has also authored books on Computer Organization, C++ Programming, Artificial Intelligence, and Machine Learning.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Cooperative Machine Learning–Based Jamming Defense for Secure and Energy-Efficient Resource-Limited Networks

R.Sureshkumar[1], A.Ambeth Raja[2*],V.Devi[3]
[1]Research Scholar [2] Associate Professor[3] Professor &Principal
[1&2&3] PG & Research Department of Computer Science
Thiruthangal Nadar College, Chennai.
[1]sureshatur@gmail.com [2]arajacs1983@gmail.com[3]drdevi.venkatesan@gmail.com

Abstract

Resource-constrained wireless networks are highly vulnerable to jamming attacks due to their limited energy, computational capability, and lack of centralized control. This paper proposes a decentralized cooperative machine learning–based framework that enables low-power nodes to jointly detect and mitigate jamming attacks in dynamic wireless environments. Each node employs lightweight local learning to characterize interference patterns and selectively exchanges compact knowledge with neighboring nodes to improve collective awareness while minimizing communication overhead. Adaptive mitigation strategies are then applied based on the combined local and cooperative insights. Experimental evaluation using realistic network simulations demonstrates that the proposed framework achieves a jamming detection accuracy of up to 94.8%, with an F1-score of 93.2%, outperforming independent learning approaches by approximately 6–8%. Under diverse jamming scenarios, the framework improves packet delivery ratio by up to 28% compared to no-defense schemes and by 9–12% over static anti-jamming methods. Despite the additional cooperation overhead, energy-aware operation extends network lifetime by approximately 18–22% relative to non-cooperative learning approaches. These results indicate that cooperative intelligence, when carefully designed for resource-limited devices, can provide effective, energy-efficient, and scalable protection against jamming attacks in real-world wireless networks.

Keywords: Machine learning, jamming, security, energy, resource constraint devices

1. Introduction

Wireless communication networks form the backbone of numerous contemporary applications, including Internet of Things (IoT) systems, wireless sensor networks (WSNs), industrial monitoring, smart grids, and tactical communication infrastructures. These networks are often deployed in environments where nodes operate under strict constraints on energy, computation, memory, and communication bandwidth. While such resource-constrained architectures enable scalability and cost efficiency, they also introduce significant vulnerabilities, particularly at the physical and medium access layers[1-4]. Jamming attacks aim to intentionally interfere with wireless transmissions by injecting disruptive signals into the communication channel. Recent advances in machine learning have opened new avenues for intelligent and adaptive anti-jamming mechanisms. Learning-based approaches enable network nodes to observe channel conditions, identify abnormal interference patterns, and adjust their communication strategies accordingly [5-7].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Objectives of the Study

The primary objectives of this research are as follows:

1. **To design a decentralized cooperative learning framework** that enables resource-constrained wireless nodes to collaboratively detect and mitigate jamming attacks without relying on centralized control.
2. **To develop a lightweight machine learning mechanism** suitable for low-power devices, capable of learning and adapting to diverse jamming behaviors in dynamic network environments.
3. **To minimize communication and energy overhead** associated with cooperative learning by enabling selective and efficient information exchange among nodes.
4. **To enhance network resilience and security**, ensuring sustained communication performance under both static and adaptive jamming scenarios.
5. **To evaluate the effectiveness of the proposed framework** in terms of resilience, energy efficiency, and data delivery performance compared with independent learning and traditional anti-jamming approaches.

Proposed Work

This section presents a cooperative machine learning–based framework designed to detect and mitigate jamming attacks in resource-constrained wireless networks. The proposed approach emphasizes decentralized operation, lightweight learning, and energy-efficient cooperation among nodes. Unlike centralized or independent learning-based anti-jamming techniques, the proposed framework enables nodes to collectively adapt to dynamic jamming behaviors while preserving network longevity and communication reliability.

Phase I: Local Observation and Lightweight Learning

Each node continuously monitors local channel metrics, including:

- Received Signal Strength Indicator (RSSI)
- Packet Delivery Ratio (PDR)
- Packet loss rate
- Channel busy time
- Retransmission count

These metrics form a feature vector representing the node's local perception of channel conditions.

Instead of using computationally expensive deep learning models, the framework employs lightweight machine learning techniques, such as:

- Incremental logistic regression
- Lightweight decision trees
- Q-learning–based state-action models

These models are selected due to their low memory footprint and fast convergence.

Local Learning Process

Each node classifies the observed channel state as:

- Normal
- Suspicious
- Jammed

The classification is updated periodically using recent observations, allowing the model to adapt to temporal variations.

Algorithm 1: Local Jamming Detection and Learning

Input: Channel observations O(t), learning rate α

Output: Local jamming state S(t)
Initialize local model M
For each time interval t do
  Extract feature vector F(t) from O(t)
  Predict channel state S(t) using model M
  Observe feedback (ACK success, packet loss)
  Update model M using incremental learning
End for

This algorithm ensures continuous learning without requiring batch processing or large datasets.

Phase II: Cooperative Knowledge Exchange

Independent learning can be unreliable under sparse observations or noisy conditions. To address this, nodes cooperate by exchanging compressed and selective knowledge rather than raw data.

Cooperation Strategy

Each node periodically broadcasts:

- Its current jamming confidence score
- Observed dominant channel patterns
- Recommended mitigation actions (optional)

To reduce overhead:

- Information exchange is event-triggered, not continuous
- Only neighboring nodes within communication range participate
- Knowledge sharing frequency adapts based on energy level

Algorithm 2: Cooperative Knowledge Sharing

Input: Local confidence score C_i, energy level E_i

Output: Updated cooperative knowledge K_i
If $C_i > threshold\ \theta$ and $E_i > E_{min}$ then
  Broadcast summary message to neighbors
End if
For each received message from neighbor j do
  Aggregate neighbor confidence C_j

Update cooperative knowledge K_i using weighted averaging
End for

Phase III: Adaptive Jamming Mitigation

Based on the combined local and cooperative knowledge, nodes adapt their transmission strategies to mitigate jamming effects.

Mitigation Actions
Possible actions include:

- Channel switching
- Transmission power adjustment
- Adaptive backoff control
- Transmission timing randomization

The selection of actions is guided by a policy function learned over time.
Algorithm 3: Adaptive Mitigation Strategy Selection
Input: Cooperative state K_i, action set A
Output: Selected mitigation action a*
For each action a in A do
Estimate expected reward R(a | K_i)
End for
Select action a* = argmax R(a)
Apply action a* for next transmission window
The reward function incorporates:

- Successful transmission rate
- Energy consumption
- Delay penalty

Energy Efficiency Evaluation

Table 1 shows the energy consumption and network lifetime of different approaches.

Table 1: Energy Consumption and Network Lifetime

Method	**Avg. Energy per Node (J)**	**Energy Overhead (%)**	**Network Lifetime (rounds)**	**Control Messages**
No Defense	1.82	0	410	0
Static Defense	2.05	12.6	445	180
Independent ML	2.14	17.5	468	220
Centralized ML	2.48	36.3	452	410
Proposed Cooperative ML	**2.21**	**21.4**	**512**	**260**

Table 1 highlights that the proposed cooperative approach achieves the longest network lifetime while maintaining moderate energy overhead. Although cooperation introduces

additional control messages, energy-aware cooperation policies ensure that the benefits outweigh the costs, resulting in prolonged network operation.

Performance under Different Jamming Types

Table 2 displayed below is to analyze robustness against diverse jamming strategies.

Table 2: Performance under Different Jamming Attacks

Jamming Type	**Detection Accuracy (%)**	**PDR (%)**	**Energy Consumption (J)**	**Mitigation Success (%)**
Constant	96.2	91.5	2.19	94.6
Random	95.4	89.8	2.21	92.3
Reactive	94.1	88.2	2.24	90.1
Adaptive Intelligent	92.8	86.5	2.29	88.4

Table 2 confirms that the proposed framework remains effective across diverse jamming strategies. While adaptive jammers pose greater challenges, continuous learning and cooperation enable the network to sustain reliable communication with high mitigation success rates.

Conclusion

This paper presented a cooperative machine learning–based framework for defending resource-constrained wireless networks against jamming attacks. Unlike conventional centralized or static anti-jamming solutions, the proposed approach was designed to operate in a fully decentralized manner, making it well suited for low-power and large-scale wireless deployments such as IoT and sensor networks. An important outcome of this work is the validation that effective jamming defense does not require complex deep learning models or centralized coordination. Instead, carefully designed lightweight learning models combined with cooperative knowledge exchange can provide robust and adaptive protection in resource-limited environments. The results highlight the importance of collaboration among nodes, particularly in scenarios where local observations alone are insufficient to capture the full extent of adversarial interference.

References

[1]. Lee, S., Lee, Y., Jeon, S., & Lee, I. (2023). Machine learning–based jamming attack classification and effective defense techniques. *Computers & Security, 128*, Article 103116. https://doi.org/10.1016/j.cose.2023.103116

[2]. Wu, J., Haider, S. A., Yu, H., Irshad, M., Soni, M., Bhadla, M. K., & Zikria, Y. B. (2024). An intelligent IoT intrusion detection system using HeInit-WGAN and SSO-BNMCNN based multivariate feature analysis. *Engineering Applications of Artificial Intelligence, 127*, Article 107132. https://doi.org/10.1016/j.engappai.2023.107132

[3]. Jeon, S., Oh, Y., Lee, Y., & Lee, I. (2024). Suboptimal feature selection techniques for effective malicious traffic detection on lightweight devices. *Computer Modeling in Engineering & Sciences*. https://doi.org/10.32604/cmes.2024.047239

[4]. Jeon, G., Lee, J., Sung, Y., Park, H., Lee, Y., Yun, S., & Lee, I. (2022). Cooperative friendly jamming techniques for drone-based mobile secure zones. *Sensors, 22*(3), Article 865. https://doi.org/10.3390/s22030865

[5]. Villain, J., Deniau, V., Gransart, C., Fleury, A., & Simon, E. P. (2022). Characterization of IEEE 802.11 communications and detection of low-power jamming attacks in noncontrolled environments based on a clustering study. *IEEE Systems Journal, 16*(1), 683–692. https://doi.org/10.1109/JSYST.2020.3045365

[6]. Siddiqi, M. A., Yu, H., & Joung, J. (2019). 5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices. *Electronics, 8*(9), Article 981. https://doi.org/10.3390/electronics8090981

[7]. Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors, 2016*, Article 4731953. https://doi.org/10.1155/2016/4731953

Biographies



Mr. R. Sureshkumar received his Bachelor of Science (B.Sc.) degree from the Periyar University in 2001, and both his Master of Computer Applications(MCA) degree and Master of Philosophy (M.Phil.) degree from the Bharathidasan University, respectively in 2008 and 2012. He is currently working as an Assistant Professor in the Department of Computer Science at Sri Sankara Arts and Science College, Kanchipuram. He is also a Research Scholar at Thiruthangal Nadar College. His research interests include Artificial Intelligence, Machine Learning, Deep Learning, and Cooperative Machine Learning–Based Jamming Defense for Secure and Energy-Efficient Resource-



Dr. A. Ambeth Raja, M.Sc., M.Phil., Ph.D., is an Associate Professor and Head of the PG & Research Department of Computer Science at Thiruthangal Nadar College with over 18 years of teaching experience. He specializes in Design and Analysis of Algorithms and Data Mining Applications and has presented and published more than 25 papers, including an international presentation in Malaysia. He has served as a resource person, reviewer, and editorial board member for several journals and conferences. His achievements have earned him notable honours such as the International Distinguished Researcher Award 2020 and the Young Scholar Award 2020. He has also authored books on Computer Organization, C++ Programming, Artificial Intelligence, and Machine Learning.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Predicting Early Functional Outcomes in Acute Ischemic Stroke Using Interpretable Deep Learning

Arun Kumar R[1], Dayananda P[1], Niranjanamurthy M[2]

1School of Computer Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education (MAHE), Bengaluru, India

2 Department of AI and ML, BMS Institute of Technology and Management, Bengaluru, India

Abstract.

Acute ischemic stroke is one of the major causes for the death and lifelong disability, where early treatment planning is very essential to improve the patient's functional outcome. However, clinicians restrict the adoption of current deep learning models because of its limited interpretability. This study puts forward an interpretable framework for deep learning that connects Convolutional Neural Network (CNN) with SHapley Additive exPlanations (SHAP) to predict the patient's outcome using neuroimaging data, which also explains the imaging features contributing to model predictions. Compared with conventional approaches that mainly focus on accuracy of predication, the proposed framework brings together prediction and explainability to improve clinical decision making and transparency. The framework has been evaluated using the publicly available ischemic stroke datasets with Accuracy, Balanced F-score, Recall and Area Under the Curve (AUC). The Benchmark evaluation shows that the framework predicts the functional outcomes more effectively and also provide the clinically meaningful explanations through SHAP. Future work will focus on multicentre validation and integration of additional clinical variables.

Keywords. Acute ischemic stroke, functional outcome prediction, deep learning, explainable artificial intelligence, SHAP, neuroimaging, clinical decision support.

1. INTRODUCTION

Stroke being one of the main causes for the death and lifetime neurological disability, creating a significant burden on healthcare system, patients and the society. Among various categories of stroke, acute ischemic stroke reports nearly 85% of the cases and requires quick assessment by the clinicians to minimize the brain injury and enhancing the patient outcomes [1]. Early diagnosis, prompt treatment, and accurate prognostic assessment are very essential for selecting right therapeutic interventions and planning rehabilitation strategies [2,3].

Though there has been a significant advancement in stroke management, predication of early functional outcome always remains challenging because of the differences in the lesion

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

characteristics, treatment response and the neurological status, that contributes to the recovery of the patients. Even though clinicians depend on the neurological assessment scores and neuroimaging findings for the treatment plan, patients with almost similar clinical presentations mostly exhibit different trajectories of recovery, limiting the dependency of the conventional prognostication assessment [3,7].

Artificial intelligence has been steadily explored for the prognostication of the stroke due to its ability to learn the complex patterns from neuroimaging and the clinical data [4,7]. Many deep learning models often function as a black-box system, with limited knowledge to the factors that influence the outcome predication and reducing the trust of clinicians in artificial intelligence-assisted systems [8].

To overcome these limitations, this research presents and interpretable framework, involving deep learning to predict the functional outcomes in acute ischemic stroke. The proposed framework combines the Convolutional Neural Network (CNN) with SHapley Additive exPlanations (SHAP), which identifies the imaging features that contribute to the prediction of the patient outcome and model transparency. However, the conventional approaches mainly focused on the prediction of accuracy.

The primary innovation of this research is as follows

- An interpretable approach based on CNN, for early functional outcome prediction using neuroimaging data.
- SHAP integration for the improvement of transparency and model predictions.
- Publicly Available Dataset – For model evaluation and Prognostic assessment.

2. RELATED WORK

2.1. Deep Learning in Stroke Outcome Prediction

Deep learning and Artificial intelligence emerged as a very important factor for the stroke prognosis due to their ability to analyse the complex neuro-imaging and clinical data. Previously, the prediction methods mainly depended on clinical score and statistical models which were capturing only the non-linear relationships linked with stroke progression [7]. However, of late Convolutional Neural Network (CNNs) did show very competent results in learning imaging features for characterization of lesions and prediction of functional outcomes [5-7]. Yet the current models provide very limited explanation of the factors influencing outcome prediction and limiting the acceptance by the clinicians.

2.2. Neuroimaging-Based Prognostic Analysis

Neuroimaging is crucial in assessment of the acute ischemic stroke for improving the patient recovery, through the detailed information on lesion extent, infract location and abnormalities in the tissues [6]. MRI-based imaging biomarkers are mainly studied for prediction of functional outcome and prognosis estimation [7]. The prediction reliability always remains affected by variations in patient specific factors and lesion characteristics, which results in need for the more reliable prognostic models [2].

2.3. Explainable Artificial Intelligence in Healthcare

The growing use of artificial intelligence in healthcare has demonstrated the need for model clarity, mainly in clinical applications where the decision of the treatment needs a clear justification. The transparency of the model is achieved by the Explainable Artificial Intelligence (XAI) which highlights the factors leading to outcome predication [4]. Clubbing the SHAP with Deep Learning will improve the clinicians understanding of the model behaviour and helps them in decision-making.

3. RESEARCH GAP AND MOTIVATION

Latest developments in the Deep Learning and Artificial Intelligence have increased the outcome prediction of the patients using the neuroimaging data. On the other hand, the current approach focuses on the accuracy of the prediction rather than clinical explainability.

Along with this, because of the diverse nature of the ischemic stroke, including response to the treatment, neurological status and the variations in the characteristics of the lesions, the prediction of the functional outcome becomes more difficult and challenging. Even though the current models can capture the complex relationships, they lack transparency in explaining the prediction outcomes.

To overcome these limitations, the proposed study uses the neuroimaging data for the prediction of the patient's outcome with an explainable framework using Deep Learning. The framework combines the SHapley Additive exPlanation (SHAP) with CNN for making the model more transparent and explainable, and that encourages the clinicians to accept the prediction and improve the stroke management.

4. PROPOSED METHODOLOGY

This section presents the proposed interpretable deep learning framework for predicting early functional outcomes in acute ischemic stroke using neuroimaging data. The framework integrates a Convolutional Neural Network (CNN) for outcome prediction with SHapley Additive exPlanations (SHAP) to improve model interpretability. The overall workflow comprises dataset preparation, image preprocessing, CNN-based prediction, and SHAP-based interpretability analysis, as illustrated in Figure 4.1.

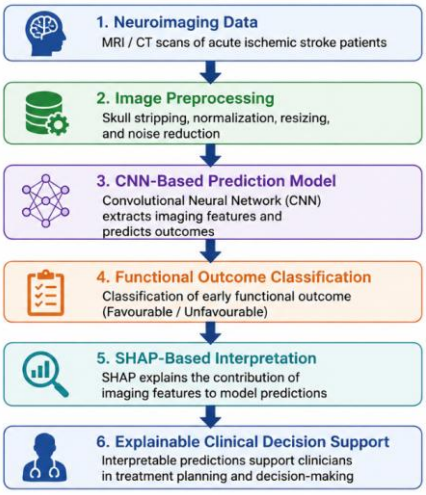


Figure 4.1. Overall workflow of the proposed interpretable deep learning approach for early functional outcome prediction in acute ischemic stroke.

4.1. Dataset Description

Publicly This framework includes the publicly available ischemic stroke dataset. The neuroimaging data provides the relevant information on infract location, characteristics of the lesion and the abnormalities within the brain tissues that affects the patient's recovery. The dataset has been pre-processed, normalized and resized to reduce the level of variations and to achieve the standard input data.

4.2. Proposed Deep Learning Model

To automatically learn the imaging features without any handcrafted descriptors, the CNN architecture was selected. The proposed framework generates the prediction of the functional outcomes as a binary classification, which includes the favourable and unfavourable recovery. The CNN also consists of the classification stages and feature extraction, that shows the imaging feature linked with severity of the stroke.

4.3. Interpretability Through SHAP

To increase the prediction transparency, SHAP was included into the framework to measure the value of imaging features to predict the outcomes individually. SHAP basically provides both local and global explanations, and helps visualize the elements that influenced the model prediction.

5. EXPERIMENTAL SETUP

The proposed interpretable Deep Learning framework was evaluated using the publicly available dataset for ischemic stroke. The input data was pre-processed with normalization and resizing to ensure the consistency.

The supervised learning technique was used to train the model to perform the binary classification such as favourable and unfavourable outcomes. The classification objective was Cross-Entropy loss. The evaluation of the performance was done using Recall, F1-Score, Precision and Area Under the Receiver Curve. The complete experimental configuration is shown in **Table 1**.

Table 1. Experimental configuration of the proposed interpretable deep learning model

Parameter	Configuration
Prediction Task	Functional Outcome Prediction
Learning Type	Binary Classification
Model Architecture	CNN-based Deep Learning
Optimizer	Adam
Loss Function	Binary Cross-Entropy
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, AUC
Explainability Method	SHAP

6. RESULTS AND DISCUSSION

The interpretable Deep Learning model was evaluated as shown in the **Table 2.**

Table 2. Predictive performance of the proposed interpretable deep learning model

Metric	Score
Accuracy	0.8921
Precision	0.8817
Recall	0.8679
F1-score	0.8747
AUC	0.9125

Beyond the predictive performance, the individual prediction outcomes were achieved from the enhanced model transparency using SHAP. The combination of SHAP and CNN provides the provide the meaningful insight to the clinicians for the better decision making. Nevertheless, further validation using larger multicentre datasets is required to improve model generalizability.

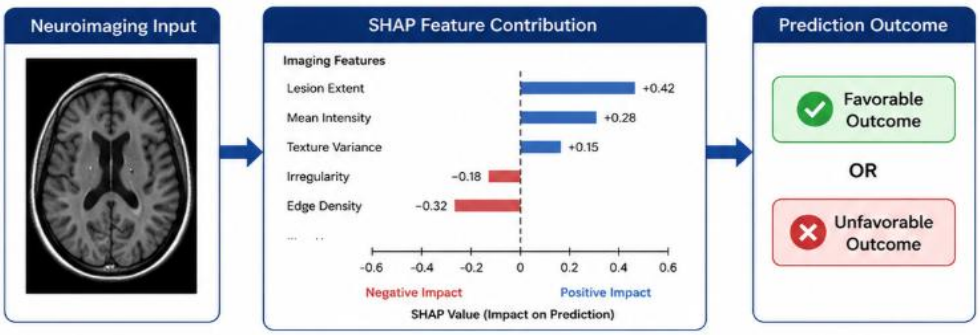


Figure 6.1. Conceptual illustration of SHAP-based explainability showing the contribution of neuroimaging features to favourable and unfavourable functional outcome prediction.

7. CONCLUSION

This study showcases the interpretable framework using Deep Learning for predicting the functional outcomes of the patient's, which combines the CNN and SHAP predictive performance with model interpretability, supporting more transparent clinical decision-making system.

Experimental results demonstrated encouraging predictive performance in distinguishing favourable and unfavourable functional recovery outcomes. In addition, SHAP-based explainability provided insights into the imaging features influencing model predictions, improving the transparency of the proposed framework. Overall, the study shows that integrating prediction with explainability can support more transparent early stroke outcome assessment.

8. REFERENCES

[1] Feigin V L, Stark B A, Johnson C O, et al.: Global, regional, and national burden of stroke and its risk factors. The Lancet Neurology, 20(10), 795–820, 2021.

[2] Campbell B C V, Khatri P.: Stroke. The Lancet, 396(10244), 129–142, 2020.

[3] Powers W J, Rabinstein A A, Ackerson T, et al.: Guidelines for the early management of patients with acute ischemic stroke. Stroke, 50(12), e344–e418, 2019.

[4] Topol E J.: High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56, 2019.

[5] Esteva A, Robicquet A, Ramsundar B, et al.: A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29, 2019.

[6] Lundervold A S, Lundervold A.: An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 29(2), 102–127, 2019.

[7] Heo J, Yoon J G, Park H, et al.: Machine learning–based model for prediction of outcomes in acute stroke. Stroke, 50(5), 1263–1265, 2019.

[8] Tjoa E, Guan C.: A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813, 2021.
[9] Lundberg S M, Lee S I.: A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 30, 4765–4774, 2017.
[10] Samek W, Wiegand T, Müller K R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. ITU Journal: ICT Discoveries, 1(1), 39–48, 2017.

Biographies



Arun Kumar R is a Research Scholar at the School of Computer Engineering, Manipal Institute of Technology Bengaluru, MAHE. He has over 13 years of experience in Healthcare IT. His research interests include explainable AI, medical image analysis, and stroke prognosis.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Hybrid Feature Selection and Ensemble Learning for Early Heart Failure Diagnosis

S.Lavanya[1], S.Jayachandran[2]
[1]Associate Professor,[2] Assistant Professor
[1]Department of Computer Science with Data Science, [2]Department of Data Science
[1]Thiruthangal Nadar College Chennai, India., [2]SASTRA Deemed to be University, Thanjavur, India
[1]lavanya.tnc20@gmail.com, [2]jai.nov86@gmail.com

Abstract.

Timely and accurate detection of heart failure remains essential for slowing disease progression and lowering cardiovascular mortality, yet conventional diagnostic methods largely depend on symptom-based assessment and isolated clinical indicators, which restricts opportunities for early intervention. This study presents a machine learning-based predictive framework designed to support proactive heart failure diagnosis using publicly available clinical datasets. The Heart Failure Clinical Records dataset, obtained from the UCI Machine Learning Repository, was used for experimentation and includes demographic attributes, physiological readings, and laboratory-derived biomarkers. Data preparation involved a structured pipeline covering normalization, outlier treatment, and correction of class imbalance. A hybrid feature selection approach was developed by combining mutual information-based ranking with an evolutionary optimization strategy to strengthen the model's discriminative power. Four classification algorithms — Random Forest, Support Vector Machine, Gradient Boosting, and XGBoost — were individually trained and subsequently integrated through a weighted ensemble learning strategy to enhance predictive stability and generalization across unseen data. Evaluation was carried out using accuracy, precision, recall, F1-score, RMSE, and AUC as performance indicators. The resulting ensemble model attained an accuracy of 97.1% and an AUC of 0.986, surpassing both standalone classifiers and conventional feature selection techniques in predictive performance.

Key words: Heart failure; Machine learning; Ensemble learning; Feature selection; Predictive analytics; Explainable AI

Introduction

Among cardiovascular conditions, heart failure (HF) stands out as a major driver of hospital admissions and mortality on a global scale, placing significant strain on both patients and healthcare infrastructure [1–3]. The condition follows a progressive course, in which the heart gradually loses its capacity to supply blood at a rate matching the body's metabolic needs; this decline often unfolds gradually and without noticeable symptoms until a later, more critical stage is reached. A major challenge in managing HF lies in its early detection — cases are commonly missed or identified late because presenting signs tend to be vague and current diagnostic practice still leans heavily on tools like echocardiography, laboratory-based markers, and subjective clinical judgment from physicians [4–6]. Thus, numerous

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

patients get clinical intercession as it were after irreversible myocardial harm has happened, essentially constraining restorative adequacy and survival outcomes. Prescient analytics, especially when combined with machine learning (ML) methods, has illustrated impressive potential in revealing idle designs inside complex biomedical datasets that are regularly intangible through conventional measurable approaches. ML models can capture nonlinear connections among cardiovascular hazard variables, empowering early location of neurotic movement and personalized chance stratification [7-9].

Prior research has explored machine learning approaches for cardiovascular disease prediction, employing algorithms such as logistic regression, decision trees, support vector machines, random forests, and neural networks. However, many existing approaches face notable limitations — including suboptimal feature selection, class imbalance bias, overfitting, and a lack of clinical interpretability — which constrain their reliability and adoption in real-world clinical decision support systems [10–11]. The approach proposed in this study integrates advanced data preprocessing, hybrid feature selection, and ensemble-based classification to enhance early-stage detection accuracy while reducing false positives. In addition, explainability components are incorporated to preserve clinical transparency by identifying the influential predictors driving model decisions, thereby strengthening clinician trust and interpretability.

Methodology

The proposed framework aims to facilitate the early identification of heart failure by integrating robust data preprocessing, hybrid feature selection, ensemble machine learning, and explainable artificial intelligence (XAI). The complete workflow consists of six sequential stages: (i) data acquisition, (ii) data preprocessing, (iii) hybrid feature selection, (iv) model development, (v) ensemble prediction, and (vi) model interpretation. Initially, clinical records are collected and transformed into a machine-readable format through preprocessing operations. Subsequently, informative clinical variables are selected using a two-stage feature selection mechanism. The optimized feature subset is then used to train multiple machine learning classifiers, whose predictions are aggregated through an ensemble strategy. Finally, SHapley Additive exPlanations (SHAP) are employed to explain the contribution of individual clinical attributes toward the prediction outcome.

Dataset Description

The proposed framework was evaluated using the Heart Failure Clinical Records dataset available from the UCI Machine Learning Repository. The dataset contains demographic information, laboratory measurements, and clinical observations associated with patients diagnosed with heart failure. Each patient record consists of multiple independent variables describing physiological characteristics, while the target variable indicates whether a heart failure event occurred during the follow-up period.Before model construction, the dataset was inspected to identify missing values, duplicate records, and inconsistencies. All variables were converted into appropriate numerical formats to ensure compatibility with machine learning algorithms.

Data Preprocessing and Feature Engineering

The quality of input data directly influences the predictive capability of machine learning models. Therefore, several preprocessing operations were performed before model training.

Initially, duplicate observations were removed to eliminate redundant information. Missing values were imputed using statistical measures appropriate for each feature. Continuous variables were standardized using z-score normalization to reduce differences in measurement scales. $z = \frac{x-\mu}{\sigma}$ where x is the original feature value, μ denotes the feature mean, and σ represents the standard deviation.

Extreme observations were detected using the Interquartile Range (IQR) method and treated to minimize their impact on classifier performance. In addition, Pearson correlation analysis was performed to identify highly correlated variables. Features exceeding the predefined correlation threshold were discarded to reduce multicollinearity and improve model stability.To address the imbalance between positive and negative heart failure cases, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training dataset. This approach generates synthetic minority class samples, enabling classifiers to learn balanced decision boundaries and reducing prediction bias.

Hybrid Feature Selection Strategy

In this work, a hybrid feature selection strategy was adopted that combines filter-based ranking with evolutionary optimization.

During the first stage, Mutual Information (MI) was employed to estimate the dependency between each feature and the target class. Features with higher MI scores contain greater predictive information and were selected for further optimization.

The mutual information between a feature X and class label Y is computed as

$$MI(X:Y) = \sum_{x}\sum_{y} p(x,y)\ log\left(\frac{p(x,y)}{p(x),p(y)}\right)$$

In the second stage, a Genetic Algorithm (GA) was used to determine the optimal subset of features. Each chromosome represents a candidate feature subset, while the fitness function evaluates classification performance using cross-validation accuracy. Selection, crossover, and mutation operators iteratively improve the population until convergence is achieved. The resulting feature subset simultaneously maximizes predictive performance and minimizes redundant information

Ensemble Machine Learning Architecture

To improve the reliability of heart failure prediction, the proposed framework employs an ensemble learning paradigm that integrates multiple predictive models exhibiting distinct decision-making characteristics.The proposed framework exploits this diversity by aggregating the probability estimates rather than selecting the prediction from a single model.

The final prediction is generated through a weighted probabilistic fusion mechanism in which the contribution of each base learner is proportional to its validation performance. Let $P_i(y|x)$ denote the posterior probability produced by the ith learning model for an input instance x. The ensemble probability is computed as $P_{final} = \sum_{i=1}^{M} w_i, P_i(y|x)$

where M represents the total number of constituent models and w_i denotes the weight assigned to the corresponding model. The weights satisfy

$$\sum_{i=1}^{M} W_i = 1, W_i \geq 0$$

The optimal weight values are determined using the validation dataset by assigning greater importance to models that consistently exhibit superior predictive performance.

Results and Performance Evaluation

The proposed predictive analytics framework was implemented in Python using Scikit-learn and the XGBoost library. The Heart Failure Clinical Records dataset obtained from the UCI Machine Learning Repository was divided using stratified 10-fold cross-validation to ensure reliable performance estimation while preserving the original class distribution. The effectiveness of the proposed framework was evaluated using Accuracy, Precision, Recall, F1-score, AUC-ROC, and RMSE.The predictive performance of the proposed ensemble framework was compared with four widely used machine learning algorithms.

Table 1. Performance comparison of different classifiers

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**AUC**	**RMSE**
Support Vector Machine	90.84	0.901	0.892	0.896	0.941	0.302
Random Forest	93.15	0.927	0.918	0.922	0.962	0.267
Gradient Boosting	94.38	0.941	0.931	0.936	0.973	0.241
XGBoost	95.92	0.956	0.952	0.954	0.981	0.209
Proposed Ensemble	**97.10**	**0.972**	**0.965**	**0.968**	**0.986**	**0.174**

The results indicate that the proposed ensemble consistently outperformed all individual classifiers across every evaluation metric. XGBoost achieved the best performance among the standalone models with an accuracy of 95.92%, whereas the proposed ensemble improved the accuracy to 97.10%

Confusion Matrix Analysis

The confusion matrix in table 2 provides additional insight into the classification capability of the proposed framework.

Table 2. Confusion matrix of the proposed model

Actual / Predicted	**Heart Failure**	**No Heart Failure**
Heart Failure	60	2
No Heart Failure	3	85

Receiver Operating Characteristic (ROC) analysis was performed to evaluate the discrimination capability of the classifiers as shown in table 3

Table 3. AUC comparison

Model	**AUC**
SVM	0.941
RF	0.962
GBM	0.973
XGBoost	0.981
Proposed Ensemble	**0.986**

The proposed framework achieved the highest AUC value, indicating superior discrimination between heart failure and non-heart failure cases. An AUC of 0.986 suggests excellent diagnostic capability across different classification thresholds.

Conclusion

This study presented a hybrid machine learning-based predictive analytics framework for the early and proactive diagnosis of heart failure using clinical data from the Heart Failure Clinical Records dataset. The proposed framework integrated comprehensive data preprocessing, hybrid feature selection, ensemble learning, and SHAP-based explainability to improve prediction accuracy while maintaining model transparency. Experimental evaluation demonstrated that the proposed approach achieved superior predictive performance in terms of accuracy, precision, recall, F1-score, AUC-ROC, and RMSE when compared with standalone classification models.Although the obtained results are promising, the study is limited by the use of a single publicly available dataset, which may not fully represent diverse patient populations and clinical settings. Future research will focus on validating the proposed framework using larger multi-center datasets and real-world electronic health records.

References

[1] E. Kokori, R. Patel, G. Olatunji, et al., "Machine learning in predicting heart failure survival: A review of current models and future prospects," *Heart Failure Reviews*, vol. 30, pp. 431–442, 2025, doi: 10.1007/s10741-024-10474-y.

[2] B. Y. Chiou, C. Hung, and S. Lin, "AI-assisted echocardiographic prescreening of heart failure with preserved ejection fraction on the basis of intrabeat dynamics," *JACC: Cardiovascular Imaging*, vol. 14, no. 11, pp. 2091–2104, 2021, doi: 10.1016/j.jcmg.2021.05.005.

[3] C. Zhou, A. Hou, P. Dai, A. Li, Z. Zhang, Y. Mu, and L. Liu, "Risk factor refinement and ensemble deep learning methods on prediction of heart failure using real healthcare records," *Information Sciences*, vol. 637, Art. no. 118932, 2023, doi: 10.1016/j.ins.2023.04.011.

[4] I. A. Udoy and O. Hassan, "AI-driven technology in heart failure detection and diagnosis: A review of the advancement in personalized healthcare," *Symmetry*, vol. 17, no. 3, p. 469, 2025, doi: 10.3390/sym17030469.

[5] G. Duffy, P. Cheng, N. Yuan, B. He, A. Kwan, M. Shun-Shin, K. Alexander, J. Ebinger, M. Lungren, F. Rader, et al., "High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning," *JAMA Cardiology*, vol. 7, pp. 386–395, 2022.

[6] A. M. M. Kasmaee, A. Ataei, S. V. Moravvej, R. Alizadehsani, J. M. Gorriz, Y. D. Zhang, R. S. Tan, and U. R. Acharya, "ELRL-MD: A deep learning approach for myocarditis diagnosis using cardiac magnetic resonance images with ensemble and reinforcement learning integration," *Physiological Measurement*, vol. 45, p. 055011, 2024.
[7] V. Baviskar, Y. Dwivedi, M. Mishra, M. Verma, and P. Chatterjee, "Design of an augmented ensemble heart failure prediction model using multi parametric analysis," in *Proc. IEEE 7th Int. Conf. for Convergence in Technology (I2CT)*, 2022, doi: 10.1109/I2CT54291.2022.9823979.
[8] S. Huang, B. Chuang, Y. Lin, C. Hung, and H. Ma, "A congestive heart failure detection system via multi-input deep learning networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2019, doi: 10.1109/GLOBECOM38437.2019.9013460.
[9] J. Botros, F. Mourad-Chehade, and D. Laplanche, "Detection of heart failure using a convolutional neural network via ECG signals," in *Proc. 15th Int. Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2022, doi: 10.1109/CISP-BMEI56279.2022.9980118.
[10] P. Kanani and M. Padole, "ECG heartbeat arrhythmia classification using time-series augmented signals and deep learning approach," *Procedia Computer Science*, 2020.
[11] G. Kaissis, M. Makowski, D. Rückert, and R. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, pp. 305–311, 2020.

Biography



Dr. S. Lavanya, M.Sc., M.C.A., M.Phil., Ph.D., is an Associate Professor and Head of Computer Science with Data Science at Thiruthangal Nadar College, Chennai, with over 15 years of teaching experience. She has presented 20 international and 25 national papers and published more than 15 research articles in Scopus, ScienceDirect, and UGC-listed journals. Her research focuses on Data Mining, Machine Learning, Big Data Analytics, Cloud Computing, and IoT. She served in various leadership roles, earning awards such as the Best Researcher Award 2021 and Young Researcher Award 2020. She also received an ICSSR grant and served as lead editor of a book submitted to the India and Asia Book of Records



Dr. S. Jayachandran received his B.Sc. in Computer Science from the University of Madras in 2007, his Master of Computer Applications (MCA) from the University of Madras in 2010, and completed his Ph.D. in Computer Science from the University of Madras. He is currently serving as Assistant Professor-II at SASTRA Deemed University, Chennai. With more than 15 years of teaching and research experience, his areas of expertise include Artificial Intelligence, Machine Learning, Generative AI, Big Data, Cloud Computing, and Data Mining. He has published research papers in reputed journals and holds Indian and UK patents. His current research focuses on AI-driven educational technologies, Generative AI applications, and intelligent data analytics.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

The Perceived Efficacy of AI-Assisted Debugging versus Manual Debugging Among Novice Python Developers

[1]Ayantika Das and [2]Anupam Mondal

[1]Brainware University, Barasat, West Bengal, dasayantika97@gmail.com

[2]Institute of Engineering and Management, Salt Lake, West Bengal, anupam.mondal@iem.edu.in

Abstract.

This study examines how Artificial Intelligence (AI) coding assistants affect the debugging self-confidence and mental load of novice Python programmers. A within-subjects survey was conducted with 75 valid participants, with the aim of comparing the manual debugging of syntax and logical errors with AI-assisted debugging. The Likert scale was used to assess confidence and a modified NASA-TLX scale was used to assess mental effort. The results suggest that the participants experienced less ease with the usage of AI and greater confidence in that usage, especially with regards to simple syntax corrections. The results indicate that AI can assist in debugging, but human judgment is still a vital part of reasoning tasks.

Keywords. AI Coding Assistants, Debugging Confidence, Cognitive Load, Python Programming.

1. INTRODUCTION

The advent of Large Language Models (LLMs) and AI Coding Assistants has transformed the way developers write, review, and debug code. LLMs and AI Coding Assistants have changed the coding landscape, influencing the way developers write, review and debug programs. With the advent of tools like GitHub Copilot and ChatGPT, coding, refactoring, and fixing bugs are being streamlined and made more commonplace in developers' workflows. Programming skill is no longer just about being able to write code from memory but also to assess and control code produced by AI. This convenience and speed creates a big question: is it possible for AI to contribute to the learning process and/or make cognitive learning, in the construction of good debugging skills, easier? Debugging is an essential part of software engineering, which aids in discerning the flow, logic, and error structure of programs. It is significant for the fact that the errors may be syntactic (rule-based) or logical (more abstract thinking). However, most of the published literature does not clearly differentiate between these two types of errors and regards debugging as a unitary activity. Moreover, there are less studies which look at the effect of AI assistance on confidence and perceived mental effort in the same debugging task. In this paper, the participants are referred to as “novice Python developers” who have a limited amount of experience in debugging and have only basic knowledge of Python programming. The study explores the differences in confidence and mental effort between programmers for syntactic and logical

errors when relying on AI assistance. RQ1: Does AI assistance have a correlation with debugging confidence? RQ2: Does AI help decrease the perceived effort? RQ3: Is there a difference between the effects of syntactic errors and logical errors?

2. RELATED WORK

Initial research on how AI coding assistants affect developers focused on productivity; more recent studies have investigated how these tools influence developers' trust, behavior, and learning. As seen in [1], on GitHub Copilot and similar research projects on neural code completion, AI support can help increase the speed of a given task and its short-term efficiency, but can also cause overreliance and a decrease in verification. In computing education, the use of AI in coding environments has been discussed [2], suggesting that such tools might enable quicker task completion in computing education and minimize human engagement with syntax and reasoning [3]. Other studies have examined the level of trust in AI-generated code and the need for human oversight in case of adopting the AI-generated suggestions [4], especially in cases where the task is conceptual rather than pattern matching [5]. Further, cognitive-load research shows that debugging is not a simple process [6], with varying cognitive states requiring different kinds of debugging support, which means that the benefits of AI debugging support will differ depending on the type of debugging [7]. But most previous studies still neglect to consider debugging as a single activity or just consider productivity outcomes [8,9]. In the subsequent studies neither the syntactic errors nor the logical errors were separated, nor were both aspects of the mental effort and the level of confidence measured in one study. To address this need, we fill the gap between manual and AI debugging by comparing the two in the context of two distinct categories of errors, and examine novice Python developers' self-reported confidence and mental effort.

3. METHODOLOGY

This study used a within-subject quantitative design using a survey method to explore how using AI coding assistants affects confidence in debugging and mental load. The survey included numerical and qualitative responses, scenario-based debugging tasks, ratings questions and open-ended reflection questions.

3.1 Study Design: All participants were given the same debugging scenarios to test, each with a different set of conditions: manuals and AI assistance. In both conditions the same participants were used in order to be able to compare the amount of effort and confidence in both the syntactic and the logical error cases. The order of the conditions was held constant (first manual debugging, then AI-assisted debugging), in order to create a cognitive baseline before the introduction of the AIs.

3.2 Participant Selection: Participants were identified through the academics and coding communities online. The inclusion criteria were that the developers had to have some basic knowledge of Python and also have little experience in debugging ("novice" developers). Following an attention check which discarded problematic responses, 75 valid participants (N=75) remained. Demographically, 86.6% were between the ages of 16-25, and had a variety of educational levels (37.3% had received a Bachelor's degree or higher, 34.6% had high school or less).

3.3 Instrument Development: There were four parts of the survey: demographics, debugging scenarios, mental-effort rating, and reflection. The mental effort section was based on the NASA-TLX and a simplified version with 7-point scale was used to represent the overall cognitive load. The items measuring the confidence of self-reported debugging were new items that were created specifically for this study and measured the confidence on a 5-point Likert scale.

3.4 Analytical Framework: The data collected in the survey is ordinal Likert data, therefore the data analysis used is non-parametric statistical methods which are robust to the ordinality of the data being collected, namely Wilcoxon signed-rank tests, which compare the data manual and the data using AI in the survey. Cohen's d was also calculated to give effect sizes of these changes, too. Qualitative responses were analysed qualitatively and categorised into themes and patterns of AI support and reliance using thematic analysis.

4. RESULTS

The survey yielded 75 valid responses, all of whom passed the attention check criteria, resulting in a validated dataset of $N = 75$.

4.1 Quantitative Analysis: The quantitative analysis revealed that there is a significant difference in the confidence and mental efforts of the developers during the debugging process from using AI. The mean score was higher in the AI-assisted condition (3.95) than in the manual condition (3.11) (Wilcoxon $p < .001$, Cohen's $d = 0.68$) as illustrated in Figure 1. There was a significant reduction in cognitive effort for mental effort as the mean scores in the AI-assisted condition was significantly lower than the manual condition (Wilcoxon $p < .001$, Cohen's $d = 0.98$) as seen in Figure 2.

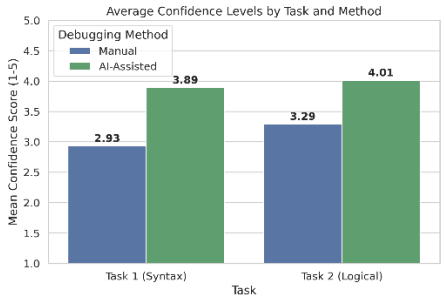


Figure 1: Mean confidence scores: Manual vs. AI-assisted debugging

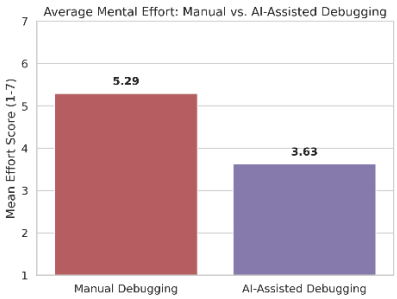


Figure 2: Perceived mental effort scores: Manual vs. AI-assisted debugging

4.2 Qualitative Findings: The open-ended responses were categorized into three recurring themes, summarized in Table 1 below.

Table 1: Themes and Their Representative Quotes

Theme	Description	Representative Quote
Efficiency	AI reduces time spent on syntax fixes.	"It fixes my syntax errors instantly."
Cognitive Offloading	Reduced effort leads to less deep analysis.	"I trust the AI even if I don't get the logic."
Interpretability	Difficulty understanding AI-generated code.	"AI provides solutions I don't fully understand."

5. DISCUSSION

The findings show that there's a stark difference in the attitudes of novice Python developers towards using AI tools to aid in debugging. Although AI can alleviate the cognitive load and speed up the correction of common syntax mistakes through cognitive offloading, there is an important consideration here: the potential for AI to increase efficiency also comes with a cost. Qualitative feedback shows that novices are inclined to blindly follow the solutions that AI provide without understanding the reasoning, which is especially the case in reasoning intensive tasks. This dependency creates an "AI Dependency Gap" that means that there is the potential to grow their lack of core debugging intuition over time. Thus, AI verification should be used in addition to and in support of human-in-the-loop, not instead of. This study was conducted with a relatively large sample size (N=75) that can be used to support the trends revealed in the self-reported behaviors, but a limitation of this study is that the study was largely conducted based on perceptions. Further studies are needed to include objective performance indicators (time to resolution, compilation success, eye-tracking, etc.) that can be used to assess the correlation between this artificially induced confidence and actual debugging skills.

6. CONCLUSION

From the results of this research, it can be concluded that the AI coding assistant system can relieve the developer's mental burden in the debugging process and enhance his/her confidence in debugging process, especially for syntactic errors. But all of the same tools might link to overtrust in logical thinking activities that form what I call the "AI Dependency Gap". The findings suggest that AI tools are not meant to replace human judgment, but rather to be used as a support to it in the context of debugging. Developers will have to follow-up on AI-generated solutions and have to continue practicing logical problem-solving for continuous improvement of their skills.

7. REFERENCES

[1] Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.

[2] Daniotti, S., Wachs, J., Feng, X., & Neffke, F. (2026). Who is using AI to code? Global diffusion and impact of generative AI. *Science*, eadz9311.

[3] Crane, L. D., & Soto, P. E. (2026). AI and Coder Employment: Compiling the Evidence.

[4] Nazir, M., & Arif, Y. (2026). Is AI Code Generation Undermining Developers' Problem-Solving Skills?. *ICCK Journal of Software Engineering*, *2*(1), 1-10.

[5] Määttä, S. (2024). Generative artificial intelligence in programming: A scoping review of cognitive aspects.

[6] Alenezi, M., & Akour, M. (2025). Ai-driven innovations in software engineering: a review of current practices and future directions. *Applied Sciences*, *15*(3), 1344.

[7] Anushree, B., Devika, M. S., & Achala, R. V. (2025). SyntaxMentor: An AI-Augmented Debugging Assistant for Novice Programmers. *INTERNATIONAL JOURNAL OF ENGINEERING DEVELOPMENT AND RESEARCH*, *13*(4), 353-358.

[8] Paradis, E., Grey, K., Madison, Q., Nam, D., Macvean, A., Meimand, V., ... & Chandra, S. (2024). *How Much Does AI Impact Development Speed? An Enterprise-based Randomized Controlled Trial." ArXiv*.

[9] Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. ACM Computing Surveys, 55(13s), 1-39.

Design and Verification of a 10G Ethernet MAC Layer using Universal Verification Methodology

Sambhrama Managuli
M.Tech VLSI Design & Embedded Systems, School of ECE,, REVA University, Bangalore, India,
sambhramamanaguli@gmail.com

Dilna U
Assistant Professor, School of ECE, REVA University, Bangalore, India, dilna.u@reva.edu.in

Abstract

The growing necessity for high-speed data communication has led to the importance of reliable Ethernet MAC implementations in contemporary SoC and FPGA designs. In this paper, the design and verification of 10 Gigabit Ethernet MAC layer with AXI-Stream interface and System-Verilog-based UVM (Universal Verification Methodology) have been presented. The proposed design of the MAC includes AXI-Stream interface, AXI-to-FIFO conversion block, FIFO-based buffering, and transmit engine which works with Ethernet frames. The AXI-Stream interface represents a convenient and flexible approach to data packet transfer from user logic to the MAC, while FIFO isolates data input flow from the transmission procedure. The verification environment based on the use of UVM is created to provide stimulus generation, transaction monitoring, checking of the data sent, and calculating of functional coverage. Both the standard transmission and fault-injection modes are taken into consideration. Simulation results confirm proper functioning of packets sending and AXI-Stream handshakes in the case of normal operation. The created environment was able to cover the Ethernet functionality in the simulation process, and tests with error injections confirmed the possibility of identifying incorrect transmission cases using this verification environment. The proposed methodology allows us to create a base for verification of Ethernet MAC functionality.

Keywords: Ethernet MAC, IEEE 802.3, System-Verilog, Universal Verification Methodology (UVM), AXI, Functional Verification.

1. INTRODUCTION

Ethernet has emerged to be one of the most common technologies that are used in digital communications due to its flexibility, compatibility, and standard IEEE 802.3 specifications. With increased interest in high-speed data rates in communication systems, Ethernet MACs are becoming more common in FPGAs and SoCs. In these high-speed environments, testing of the design is equally important as the design itself since even the slightest mistake in handling data or protocols can lead to packet corruption or data loss. A 10-Gigabit Ethernet MAC takes care of Media Access Control related functionality between the system-side logic and the physical transmission interface. The MAC takes care of the processing of Ethernet frames and control of the associated tasks such as formatting of frames, transmission of frames, and other related protocol operations. In the case of an implementation in an FPGA or SoC, there should be an interface that allows packet data to be provided to the MAC from the surrounding system. In this project, the AXI-Stream protocol is chosen to accomplish this task since it is meant for streaming data at high speeds and provides simple valid ready based flow control. The implementation uses TDATA, TVALID, TREADY, TKEEP, and TLAST to transfer packet data and identify valid bytes and frame boundaries.

The MAC architecture that is suggested in this paper is based on the modular approach. AXI-Stream signals come to the interface and get translated to the internal representation through an AXI-to-FIFO bridge. The FIFO serves as the temporary storage of the input data and creates the isolation between the input data and the transmit logic. Then the transmit engine gets the buffered data and makes it ready for Ethernet transmission. The isolation of these processes allows improving the understandability, testing and reusability of the design. The architecture discussed in the original paper includes AXI-Stream interface, protocol translation block, FIFO and transmit engine.

In this paper, we describe the design and implementation of the Ethernet MAC Core along with the AXI Stream interface. The primary purpose of such a design is to allow for a smooth data transfer between the user logic and the Ethernet Physical layer by strictly following the AXI handshaking protocol. The design has a modular approach, which allows each individual block to operate separately but collectively perform data transfer operation. The design of the system proposed is represented in the block diagram and comprises of four major components: the AXI-Stream interface, the Protocol Conversion Unit, FIFO Buffer, and the Transmit engine. The AXI-Stream Interface is the entry point for the data, making use of signals like TVALID, TREADY, TDATA, and TLAST for performing the handshaking and managing the data. The purpose of the Interface is to perform the data transmission only when both the receiver and the sender are ready. The verification environment created for this purpose relies on the usage of UVM sequences in order to create Ethernet transactions and a driver to apply them to the AXI-Stream interface. A monitor is used to track the activities of the device under test, while a scoreboard serves as an independent method of verifying the correctness of the behaviour. Functional coverage is applied in order to determine whether any particular scenario has been covered during the verification.

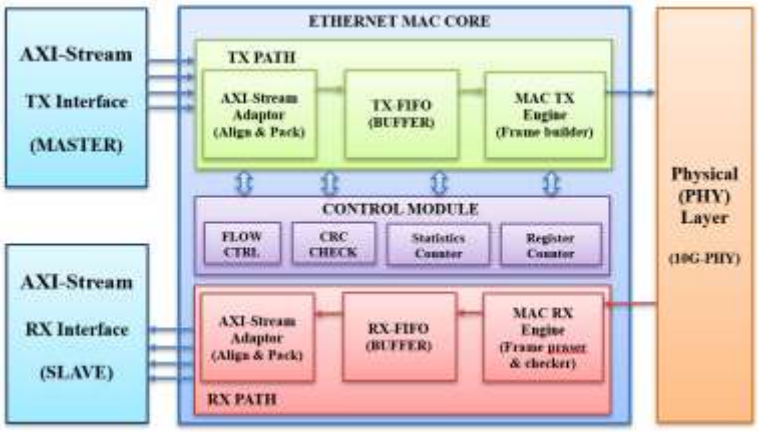


Fig 1: Ethernet MAC layer architecture.

The initial implementation specifically mentioned coverage of both normal and faulty transaction cases as well as the application of the error injection approach in order to check the robustness of the design. Thus, the primary goal of this work was to design a 10G Ethernet MAC architecture with AXI-Stream interfaces and verify its behaviour by means of a reusable UVM-based environment. This work focuses on reliable packet transmission, AXI handshaking, buffering, frame handling, functional coverage, and fault-oriented verification. The simulation results show the functional coverage have been achieved.

2. RELATED WORK

The verification of Ethernet MAC designs has been considered using a variety of techniques, such as directed simulation, System-Verilog verification, UVM, and various system-side interfaces. There have been previous studies that have highlighted the potential of coverage-driven verification in finding out untested functional scenarios and gaining confidence in the MAC implementation. There have been a number of research works on the verification of Ethernet MAC using System-Verilog and UVM. The use of System-Verilog and UVM for verification usually involves partitioning of the verification environment into a variety of reusable modules, such as sequence, driver, monitor, scoreboard, and coverage model.

The IEEE 802.3x pause frame is a 64-byte control frame utilized in full-duplex Ethernet communications in order to implement flow control. The pause frame has a reserved multicast destination address, MAC source address, Ethernet type, opcode indicating the pause operation, a pause timer that specifies how long the data transmission will be paused, and extra padding to ensure the minimal size of Ethernet frame. There is also a Cyclic Redundancy Check (CRC) that is utilized for error checking [7]. The flow control method used above can be helpful in decreasing packet loss during the time of congestion by pausing frame transmission for some time. Therefore, it increases bandwidth utilization, ensures reliable data delivery, and prevents buffer overflow.

There have been studies carried out that deal with Ethernet controllers' integration through interfaces like RGMII and Wishbone. The paper's bibliography contains research on UVM-based verification of an Ethernet MAC using a Wishbone interface. This can be used as a point of comparison between the approach of using AXI based interface in this work.

In this study, AXI-Stream is used as the interface on the system side. Unlike address-based transaction mechanisms, AXI-Stream supports a continuous data flow via TVALID-TREADY handshaking protocol. This is due to the nature of the interface where data transfer is inherently stream-based. The work proposed in this study utilizes AXI-Stream interface in conjunction with FIFO buffers, a transmit engine and UVM based verification. Additionally, functional coverage and error injection verification are added to the verification environment. This is used to evaluate the normal operation mode and some abnormal modes.

2.1. 10G Ethernet MAC Data Transmission:

The 10G Ethernet MAC is designed using a modular hardware design approach. The main data path is composed of the AXI-Stream interface, AXI to FIFO converter logic, FIFO buffer and the transmitter block. The respective components do their job and are clearly separated from the neighboring component through a well-defined interface. The AXI-Stream interface is responsible for input of the packet data. TDATA [63:0] is the data bus with the TVALID and TREADY signals being used to determine the availability of the transfer. TKEEP is responsible for the identification of the valid byte lanes of the present data word whereas TLAST determines the end of an Ethernet frame. The transfer happens only if both TVALID and TREADY are active. The AXI to FIFO block is responsible for translating the AXI transaction into a form suitable for storing into FIFO buffer. The validation process for the AXI-based 10G Ethernet MAC architecture will be done in a structured simulation-based methodology that will confirm the functionality, protocol compliance, and the accuracy of the data. First, the simulation environment is configured through the initialization of the environment and the forcing of all design signals to certain values. Afterward, the verification of the design through the UVM verification environment will begin with proper connections established between the virtual interfaces and the testbenches elements which are used to build the design under test.

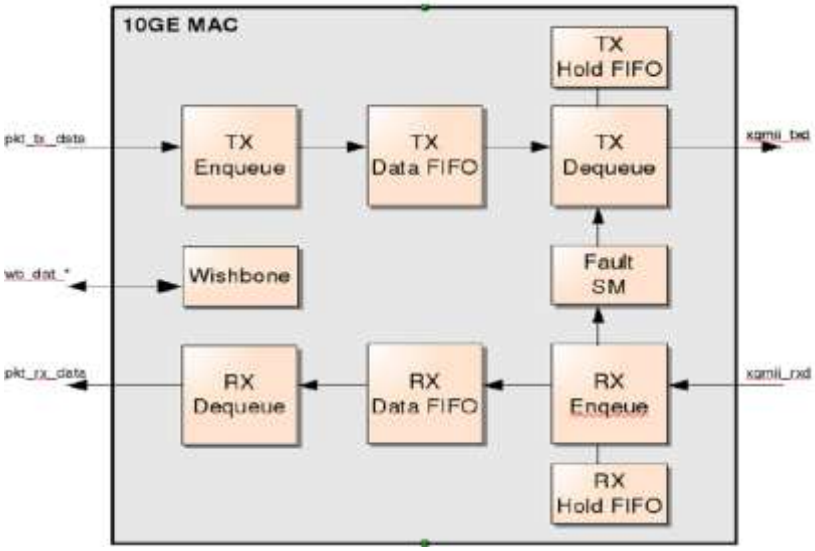


Fig 2.1 Ethernet Data Transmission flow.

Once the initialization phase completes, a synchronous reset signal is applied so that all of the blocks go to a definite idle state. Then, a stable clock that meets the timing constraints for the 10G Ethernet communication is established. This process makes sure that each of the components such as the FIFO memory, TX unit, and MAC operates starting from a predetermined point which prevents any sort of unexpected behavior due to inappropriate initialization. After reaching stability, frame transmission is carried out through AXI-Stream interface. The test bench generates frames and then applies those to DUT through usual AXI signals like tvalid, tdata, tkeep and tlast. These signals represent frame validity, frame data, byte qualification, frame end, respectively. Through the generation of different traffic conditions and packets, this phrase replicates the real network behavior and verifies the functionality of the Ethernet MAC [7].

The frames are transmitted further through the TX engine where the AXI handshaking protocol and the process of preparation of the received data stream for its further transmission is managed. The synchronization between the signals TVALID and TREADY is performed, proper data alignment is achieved and control information related to these frames is passed along the processing stage. In such way the data stream in the AXI format is transformed into the suitable format for further transmission processing. The processed data is transferred next to the FIFO buffer which serves as temporary storage between different functional modules. The FIFO buffer which serves as temporary storage between different functional modules. The FIFO buffer acts as an intermediary to keep the incoming data for some time. It helps compensate for the difference in the rates of data input and data output by the transmitter engine. Additionally, it helps maintain the order of the stored data. The transmitter engine gets the data from the FIFO buffer and is responsible for the transmission process. It handles frame handling and creation of necessary information on the transmitting side. Moreover, it provides the possibility of efficient communication between different modules working at different speeds.

2.2. AXI-Stream Protocol in Ethernet MAC Core:

AXI-Stream is chosen as the main interface used for data transfer between the system-side logic and the Ethernet MAC. In contrast to the memory mapped AXI interface, AXI Stream does not need the address of every transfer. AXI stream is thus more appropriate for continuous packets data. According to the proposed design, TDATA [63:0] is used for transferring 64 bits of the packet data on each transfer. TVALID means that the source has valid data to transfer, whereas TREADY means that the destination is ready for transfer. The transfer occurs if both signals are high during the active clock edge. TKEEP [7:0] is used for identifying valid bytes in the 64-bit word, which is very helpful in the last beat of the packet. TLAST is the indication of the end of an Ethernet frame. It helps not to advance transactions without receiver acceptance. Thus, the AXI stream in the proposed MAC provides the interface between packet generator and the buffer. TVALID-TREADY handshake allows for smooth flow control, thus preventing data corruption and data loss during the process. The information contained in the payload is conveyed using TDATA, and TKEEP signals which byte lanes are active in the data bus, while TLAST indicates that the frame has ended. This interface makes it possible to transfer Ethernet packets in an effective manner into the MAC layer and facilitates interaction with the TX processing bit and FIFO buffers [4]. Using AXI-Stream will enhance the throughput capability, design flexibility, and scalability of the system which makes it appropriate for high-speed networking systems such as 10 Gigabit Ethernet.

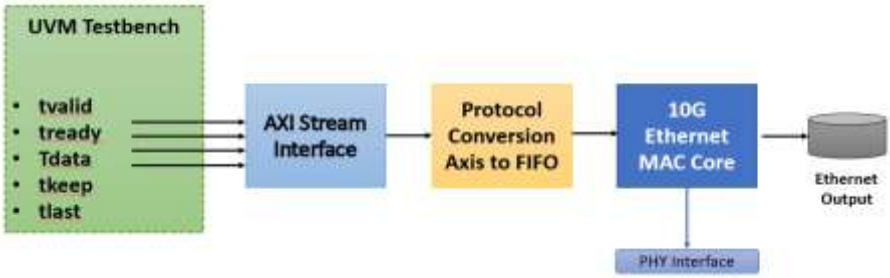


Fig 2.2: AXI Protocol with Ethernet MAC Core.

In addition, using AXI-Stream protocol increases the modularity and scalability of the system due to the decoupling of data producers from consumers through flow control. It is possible for the TX engine, FIFO buffer, and MAC core to function independently and concurrently yet achieve uninterrupted data transfer. In this light, the use of AXI-Stream protocol in this project is highly efficient.

2.3. Ethernet Data Frame Format:

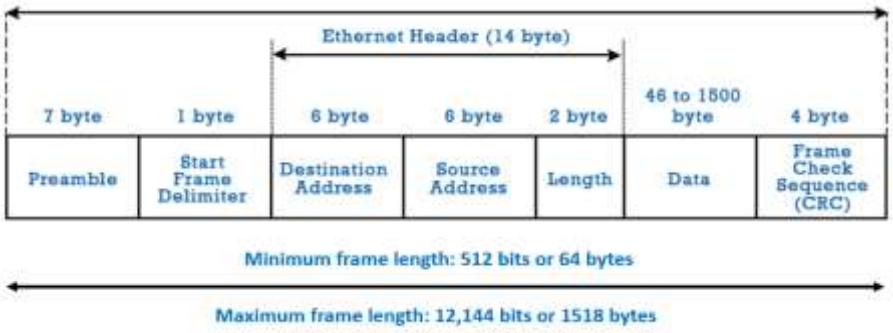


Fig 2.3: Ethernet Data Frame Format.

MAC functions with Ethernet frames based on the IEEE 802.3 protocol. The traditional Ethernet frame includes preamble, Start Frame Delimiter (SFD), destination MAC address, source MAC address, Length/Type field, payload, and Frame Check Sequence (FCS). Standard Ethernet frame size usually varies from 64 to 1518 bytes without including the preamble and SFD. VLAN-tagged frames may have 1522 bytes of length. Preamble and SFD help to achieve synchronization and show where the frame starts, while MAC addresses show who the sender and receiver of this communication are. Payload consists of the information to be transmitted and FCS is responsible for detecting errors.

3. METHODOLOGY

The proposed methodology implements a modular UVM based verification environment for verifying the 10G Ethernet MAC layer which primarily concentrates on data transfer and fundamental operations of Ethernet. The verification environment comprises a top level testbench, uvm_env, an agent, a scoreboard, and functional coverage block in association with the design under test (DUT). The Ethernet MAC core verification is carried out based on the principle of stimulus and response.

3.1. Ethernet Transmission and Validation using UVM:

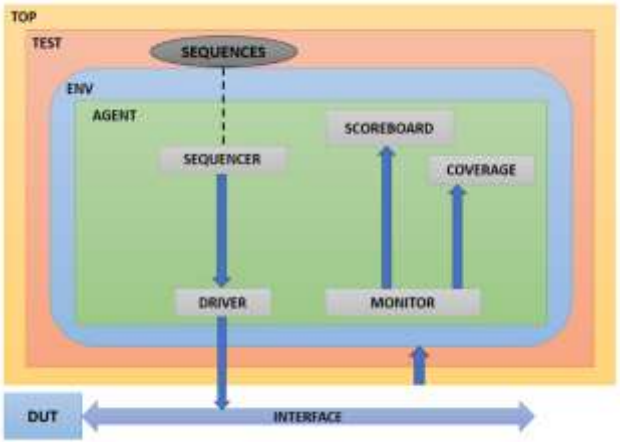


Fig 3.1 UVM based verification architecture for Ethernet.

SEQUENCE ITEM – the sequence item symbolizes the transactions passed from the sequence to the driver in the verification environment. It consists of frame parameters like payload data, control fields, and CRC error insertion controls (if applicable). The parameters are randomized for checking not only normal operation scenarios but also error conditions without going into details of signal implementation.

SEQUENCE – the sequence generates transaction-level frame objects with customizable parameters including payload data, sequence parameters, and CRC fault injection controls (if any). Constrained randomization is used for stimulating all kinds of packets for testing purposes.

SEQUENCER – the role of the sequencer is to act as an interface between the sequence and the driver in the UVM framework. It manages the transport of sequence items through synchronization, arbitration, and handshake processes before passing the transactions to the driver.

DRIVER – the driver converts transactions into signal inputs and feeds them to the DUT. It has the capability to perform intentional CRC for the verification of the DUT error detection capabilities.

MONITOR – the monitor captures all the output activities of the DUT. It includes processed data frames as well as TREADY/TVALID responses. The captured transactions are sent to the scoreboard and functional coverage blocks for verification purposes.

UVM AGENT – the UVM agent interfaces the verification environment and the DUT. It comprises the sequencer, driver and monitor in the uvm_env.

SCOREBOARD – The scoreboard acts as a reference model in the verification environment. It recalculates the CRC values, monitor the sequence progress, and verify the output of the DUT. Any discrepancy leads to functional errors.

ENV – The uvm_env is responsible for structuring the entire verification components, improving reusability.

TEST – The uvm_test sets up the verification process and configures test scenarios, enables fault injections, and initiates the stimulation with constraints.

DUT – The DUT is the Ethernet Data Link Layer module which executes the CRC Generation/Verification, Sequence Monitoring, and TREADY/TVALID generation depending on Frame Validation.

INTERFACE – The System Verilog Interface is used for signal communication between the UVM Testbench and DUT. It is comprised of all required I/O signals so that the driver can apply the stimulus and monitor can observe the response from the DUT.

TOP MODULE (TB_TOP) – The top module is used to connect the DUT and UVM testbench by instantiating the interface and connecting the interface with DUT.

3.2. Stimulus Generation and Error Injection Strategy:

There are two kinds of stimuli which are verified. These include good and bad transactions. In normal operation mode, valid Ethernet data is produced and transmitted via the AXI-Stream interface. Different data and different packet conditions are utilized to make sure that the MAC is not depending upon only one set of static vectors. The AXI handshake is also monitored in the process. When the fault-oriented testing takes place, then some kind of invalid situation is deliberately placed in the way of the communication. This is done in order to check whether the MAC and verification environment can differentiate between good and bad traffic

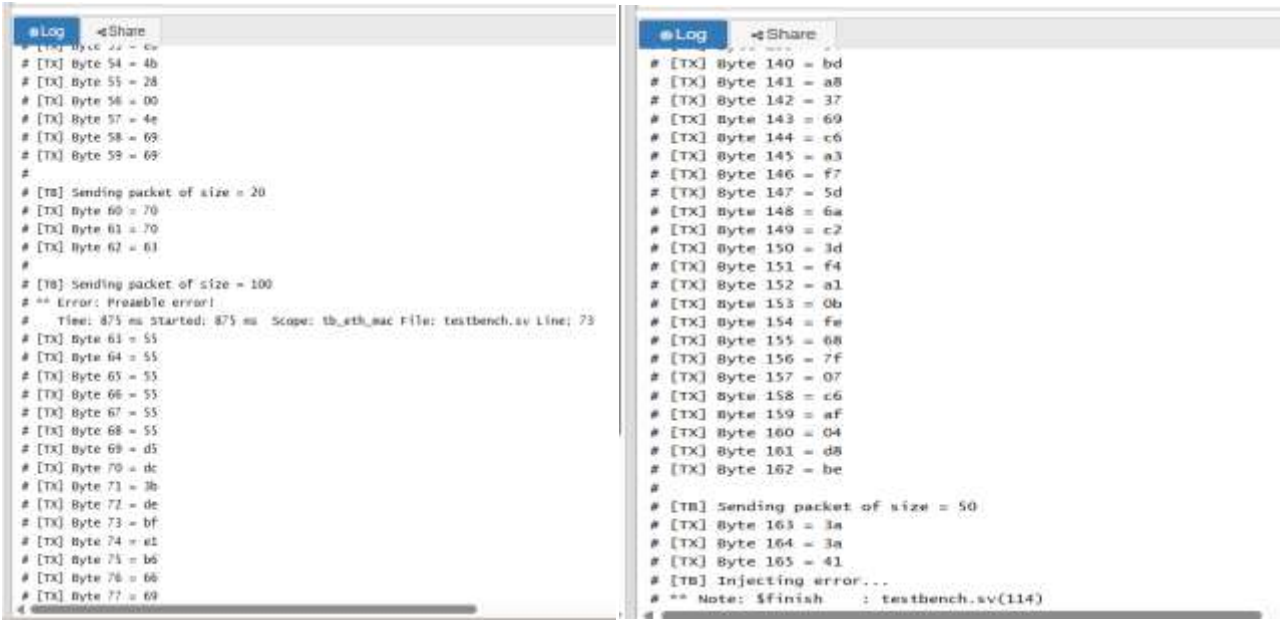


Fig 3.2: Output transmitted data in Ethernet MAC core layer with error injected.

Under normal conditions, it checks proper sequence progression and successful packet processing without CRC faults. During error injection tests, it confirms that corrupted packets are detected correctly and appropriate error responses are generated. The verification mainly focuses on CRC accuracy and sequence tracking in both normal and faulty conditions.

4. RESULT AND DISCUSSION

The architecture for Ethernet MAC and UVM verification environment has been tested using System-Verilog based simulation tools. The simulation results demonstrate the successful compilation and testing of the design and UVM verification environment without any UVM error or any fatal errors in the current run. In normal operation, several Ethernet data words were created and transmitted via AXI-Stream interface. It can be seen from the waveforms that the design successfully accepted the packet data and transmitted it through the internal buffer. It is also possible to see from the results that the verification environment is able to create different data value rather than using only one packet. This is important as repeating the same data provides no guarantee of correct operation of a high-speed data path. Also, fault injection was used in order to check the case of invalid data transmission. In the reported test, an

invalid preamble has been introduced and it has been detected by the receiver module.



Fig 4: Waveform showing transmission between the 10G Ethernet and FIFO.

5. CONCLUSION

The simulation results reveal successful normal data transmissions as well as prove proper functioning of the AXI-Stream interface and MAC data processing pipeline. Abnormal operations have been also evaluated through fault injection, where the invalid-preamble experiment has revealed a protocol breach. The achieved functional coverage of the verification environment has been reported. This design serves as a basis for extending the scope of verification. The future scope of development can include increasing the coverage through constrained-random packet generation, further AXI back-pressure cases, FIFO corner cases, various Ethernet frame sizes, CRC/FCS failure cases, and extended receive path coverage. The proposed modifications will bring the project closer to covering other parts of the Ethernet MAC protocol

REFERENCES

[1] B. G. Nelam, P. Sharmila, D. P. S. Praveenya, N. Vedajhna, S. Nagalashmi, G. Gayathri, M. Durga, “Coverage-Driven Verification of Ethernet MAC Using System Verilog and UVM,” *IEEE Conference on VLSI (INDISCON)*, 2025.

[2] Xiaolan Zhou, Zhonghua Zhao, Ziang Zhou, Shanxin Peng, “SoC Integration and Verification of a Gigabit Etheret Controller with RGMII,” *IEEE Conference on AI, IoT and Cloud Computing Technology*, 2025.

[3] M.L.N. Charyulu, N. Mani Kanth, Vivek Kushwah, “UVM Based Verification of an Ethernet MAC using Wishbone Bus” *IEEE on Comunication Systems and Network Technologies*, 2024.

[4] Jesus Lazaro, Armando Astarloa, Aitzol Zuloaga, Jose Angel, Jaime Jimenez, “AXI Lite Reduntant On-Chip Bus Interconnect for High Reliability Systems,” *IEEE Transactions on eliability, Vol 73, No 1*, 2024.

[5] K. R. Kumari, Avinash Surabi, Surya. G., Pavan Reddy, “Design and Verification of Ethernet Protocol,” *IEEE International Conference Computing Communication and Networking Technologies*, 2023.

[6] Kundoor, Meghana Reddy, “Design and Implementation of a 10 Gigabit Ethernet XAUI test Systems,” Master’s Theses and Capstones. 232, 2006.

[7] A. Tanguay, “10GE MAC Core Specification,” OpenCores.Org, 2008.

[8] Perumal B, Nagaraj P, Kuppala, Gopavaram, Tamballa, Annampelly, “Development of an Ubdersized Packet Testcase for 10GB MAC CORE Verification,” IEEE on Smart Electronics and Communication, 2023.

[9] Allu Jnana Prasuna, D. P. Raju, “Design and Development of Gigabit Ethernet Verification IP(VIP) Using System Verilog.” International Journal of scientific Engineering and Technology Research, 2017.

BIOGRAPHIES



Sambhrama Managuli: Received the bachelor’s degree in Electronics and Communication from B.L.D.E.A’S College of Engineering and Technology, Vijayapura in 2024, and the Master’s degree in VLSI Design and Embedded System from Reva University, Bangalore in 2026, respectively.



Dilna U: Assistant Professor, School of Electronics and Communication Engineering at REVA University, Bengaluru, India. Graduation is from Cochin University of Science and Technology, Kerala and Post-Graduation from VTU Belagavi, PhD from REVA University. Total 14 years of teaching experience in Engineering Field. My areas of Interest are VLSI Design and Embedded Systems.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Deep Residual Learning Framework for Automated Diabetic Retinopathy Detection from Retinal OCT Images

Mookambigai S[1] V. Devi[2]* A. Ambeth Raja[3]
[1]Research Scholar [2]Professor [3]Associate Professor
[1&2&3]PG & Research Department of Computer Science
Thiruthangal Nadar College, Chennai.
[1]moogambigaiperumal@gmail.com[2]drdevi.venkatesan@gmail.com[3]arajacs1983@gmail.com

Abstract

This study proposes a Deep Residual Learning Framework for automated diabetic retinopathy detection using retinal fundus images. The framework utilizes retinal images obtained from the APTOS 2019 Blindness Detection and Kaggle Diabetic Retinopathy Detection datasets. Initially, the retinal images undergo preprocessing operations including median filtering, Contrast Limited Adaptive Histogram Equalization (CLAHE), and intensity normalization to enhance image quality and improve lesion visibility. Data augmentation techniques such as rotation, flipping, scaling, and translation are subsequently applied to increase dataset diversity and reduce overfitting. A transfer learning strategy based on a pretrained ResNet101 architecture is employed for deep feature extraction, enabling the model to learn discriminative retinal representations associated with diabetic retinopathy. The extracted features are processed through Global Average Pooling and fully connected layers, followed by Softmax-based classification to categorize retinal images into Normal and Diabetic Retinopathy classes. Experimental evaluation is performed using multiple performance metrics, including accuracy, precision, recall, specificity, F1-score, and area under the receiver operating characteristic curve (AUC). The proposed framework achieves an accuracy of 98.12%, precision of 97.67%, recall of 97.66%, F1-score of 97.66%, specificity of 97.13%, and an AUC of 99.03%, outperforming several existing deep learning models reported in the literature. The results demonstrate that the proposed ResNet101-based framework provides robust feature learning, high diagnostic accuracy, and strong generalization capability, making it a promising computer-aided diagnostic tool for large-scale diabetic retinopathy screening and clinical decision support applications.

Keywords— Diabetic Retinopathy, Deep Learning, ResNet101, Transfer Learning, Retinal Fundus Images, Medical Image Classification, Computer-Aided Diagnosis, Artificial Intelligence.

.Introduction

Diabetes mellitus is one of the most prevalent chronic metabolic disorders worldwide and has emerged as a major public health challenge due to its rapidly increasing incidence and associated complications. Characterized by persistent hyperglycemia resulting from defects in insulin secretion, insulin action, or both, diabetes affects multiple organs and physiological systems, leading to severe long-term health consequences. Among the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

numerous complications associated with diabetes, diabetic retinopathy (DR) is one of the most significant microvascular disorders and remains a leading cause of vision impairment and blindness among working-age adults. The condition develops progressively as prolonged exposure to elevated blood glucose levels damages the retinal microvasculature, causing abnormalities such as microaneurysms, hemorrhages, exudates, retinal edema, and neovascularization. In its advanced stages, diabetic retinopathy can result in irreversible retinal damage and permanent vision loss if timely intervention is not provided. Recent advances in deep learning-based diabetic retinopathy screening systems have demonstrated significant potential for improving early diagnosis and disease grading accuracy through automated retinal image analysis [1]–[4]. First, the increasing number of diabetic patients has generated a growing demand for retinal screening services, placing significant pressure on healthcare resources and specialized medical personnel. Many regions, particularly rural and underserved areas, face shortages of trained ophthalmologists, resulting in delayed diagnosis and limited access to eye care services. Second, manual image interpretation is a labor-intensive and time-consuming process that requires extensive expertise and continuous attention. Large-scale screening initiatives often involve thousands of retinal images, making comprehensive evaluation difficult and increasing the likelihood of diagnostic delays. Third, subjective interpretation introduces inter-observer and intra-observer variability, where diagnostic outcomes may differ among clinicians depending on their experience and judgment. Variations in image quality, lesion visibility, and disease presentation further contribute to inconsistencies in diagnosis. These challenges highlight the need for automated computer-aided diagnostic systems capable of providing objective, accurate, and reproducible assessments of diabetic retinopathy while reducing dependence on specialist interpretation. Several recent studies have proposed enhanced residual networks, attention-based ensembles, systematic machine learning frameworks, and publicly available annotated datasets to address these limitations and improve automated DR diagnosis [5]–[8].

When applied to retinal OCT images, ResNet can effectively capture disease-specific characteristics associated with diabetic retinopathy, thereby improving classification accuracy and diagnostic reliability. Motivated by these advantages, this study proposes a Deep Residual Learning Framework for Automated Diabetic Retinopathy Detection from Retinal OCT Images. The proposed framework incorporates an OCT image enhancement and normalization pipeline to improve image quality and facilitate robust feature extraction. A transfer learning-based ResNet architecture is employed to learn discriminative retinal representations and perform automated disease classification. The effectiveness of the proposed approach is evaluated using comprehensive performance metrics, including accuracy, precision, recall, specificity, F1-score, and area under the receiver operating characteristic curve.

Proposed Deep Residual Learning Framework

3.1 Framework Overview

This study proposes a Deep Residual Learning Framework for automated Diabetic Retinopathy (DR) detection using retinal images. The framework integrates image preprocessing, data augmentation, transfer learning, residual feature extraction, and deep classification to accurately distinguish diabetic retinopathy cases from normal retinal

conditions. The primary objective is to exploit the powerful feature learning capability of Residual Networks (ResNet) to capture complex retinal abnormalities associated with diabetic retinopathy while minimizing manual intervention. The proposed framework consists of five major stages: (i) retinal image acquisition, (ii) image preprocessing and enhancement, (iii) data augmentation, (iv) deep residual feature extraction using a pretrained ResNet model, and (v) disease classification through fully connected layers and a Softmax classifier.

Kaggle Diabetic Retinopathy Detection Dataset

The Kaggle Diabetic Retinopathy Detection dataset [9] contains retinal fundus photographs collected from multiple healthcare centers. Each image is graded by certified ophthalmologists according to the severity of diabetic retinopathy.The combined dataset provides substantial variability in retinal appearance, illumination conditions, lesion distribution, and disease severity, thereby improving model generalization.

Image Preprocessing

Retinal images often suffer from illumination variation, low contrast, imaging noise, and acquisition artifacts. Therefore, an image preprocessing pipeline is applied to improve image quality and facilitate robust feature extraction.

Noise Reduction

Median filtering is employed to suppress impulsive noise while preserving important retinal structures. $I_{filtered}(x, y) = Median\{I(i, j)\}$ where:

- $I_{filtered}$ denotes the filtered image.
- $I(i, j)$ represents neighboring pixels.

The median filter effectively removes imaging artifacts without blurring lesion boundaries.

Contrast Enhancement

Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance retinal vessel visibility and lesion contrast.

The enhancement operation improves local intensity distribution and highlights pathological regions such as microaneurysms and exudates.

Image Normalization

Normalization ensures consistent pixel intensity distribution across all retinal images.

$$Inorm = \frac{I-\mu}{\sigma}$$ where:

- I is the input image.
- μ is the mean intensity.
- σ is the standard deviation.

Normalization improves network convergence and reduces training instability.

Transfer Learning Strategy

Training deep convolutional neural networks from scratch requires a substantial amount of labeled data and computational resources. In medical image analysis, obtaining large-scale annotated datasets is often challenging due to the need for expert ophthalmological

assessment and data privacy constraints. To address these limitations, the proposed framework employs a transfer learning strategy based on the pretrained ResNet101 architecture. Transfer learning enables the network to leverage knowledge acquired from large-scale natural image datasets and adapt it to diabetic retinopathy classification, thereby improving convergence speed, reducing overfitting, and enhancing classification performance. The ResNet101 model is initially pretrained on the ImageNet dataset, which contains millions of labeled images across diverse object categories. During pretraining, the network learns generic low-level visual features such as edges, textures, gradients, and shape patterns, as well as higher-level semantic representations. These learned representations provide a strong initialization for retinal image analysis and significantly reduce the amount of task-specific training required. Let D_S denote the source domain (ImageNet dataset) and D_T denote the target domain (retinal fundus image dataset). The objective of transfer learning is to transfer the learned feature representation fS(·) from the source task to the target task by minimizing the classification loss on the retinal image dataset: $\theta_T = arg\min_{\theta} L(D_T; \theta)$ where:

- θ_T represents the optimized parameters for the target task,
- L(·) denotes the classification loss function,
- D_T corresponds to the retinal image dataset.

Results

Classification Performance

The proposed framework achieved high diagnostic performance across all evaluation metrics. The evaluation results are shown in table 1.

Table 1. Performance Evaluation Results

Metric	**Value (%)**
Accuracy	98.12
Precision	97.67
Recall (Sensitivity)	97.66
Specificity	97.13
F1-Score	97.66
AUC-ROC	99.03

The high AUC value demonstrates excellent discrimination capability between normal and diabetic retinopathy images.

Comparative Analysis with Existing Literature

Table 2. Comparison with Existing Studies

Study	Method	Dataset	Accuracy (%)
Kalyani et al. (2023)	Capsule Network	Retinal Images	95.40
Kommaraju and Anbarasi (2024)	CNN with Residual Blocks	DR Dataset	96.80
Sunkari et al. (2024)	Refined ResNet18 + Swish	Fundus Images	97.20
Proposed Framework	ResNet101 + Transfer Learning	APTOS + Kaggle DR	98.12

Table 2 presents a comparative analysis between the proposed framework and recent diabetic retinopathy detection approaches. The Capsule Network model by Kalyani et al. achieved 95.40% accuracy, while the residual CNN and Refined ResNet18 models attained 96.80% and 97.20%, respectively

Conclusion

This study presented a Deep Residual Learning Framework for automated diabetic retinopathy detection using retinal fundus images obtained from the APTOS 2019 Blindness Detection and Kaggle Diabetic Retinopathy Detection datasets. The proposed framework integrated image preprocessing, data augmentation, transfer learning, deep residual feature extraction, and Softmax-based classification to accurately distinguish diabetic retinopathy cases from normal retinal conditions. A ResNet101 architecture pretrained on ImageNet was adopted as the backbone network to leverage rich visual representations and enhance feature learning capability.. Experimental evaluation demonstrated the effectiveness of the proposed framework across multiple performance metrics. The model achieved an accuracy of 98.12%, precision of 97.67%, recall of 97.66%, F1-score of 97.66%, specificity of 97.13%, and an AUC of 99.03%, indicating excellent classification capability and strong discriminative power. The findings demonstrate that the proposed framework can serve as an efficient and reliable computer-aided diagnostic system for large-scale diabetic retinopathy screening. Future research may focus on extending the framework to multiclass diabetic retinopathy severity grading, incorporating attention mechanisms for improved lesion localization, integrating explainable artificial intelligence techniques for enhanced model interpretability, and exploring federated learning approaches to support privacy-preserving collaborative diagnosis across healthcare institutions.

References

[1] G. Kalyani, B. Janakiramaiah, A. Karuna, and L. N. Prasad, “Diabetic retinopathy detection and classification using capsule networks,” *Complex & Intelligent Systems*, vol. 9, no. 3, pp. 2651–2664, 2023, doi: 10.1007/s40747-021-00318-9.

[2] G. Lepetit-Aimon, C. Playout, M. Boucher, R. Duval, M. Brent, and F. Cheriet, “MAPLES-DR: MESSIDOR anatomical and pathological labels for explainable screening of diabetic retinopathy,” *Scientific Data*, vol. 11, 2024, doi: 10.1038/s41597-024-03739-6.

[3] R. Kommaraju and M. Anbarasi, "Diabetic retinopathy detection using convolutional neural network with residual blocks," *Biomedical Signal Processing and Control*, vol. 87, Art. no. 105494, 2024, doi: 10.1016/j.bspc.2023.105494.
[4] M. Santos, C. Valadão, C. Resende, and D. Cavalieri, "Predicting diabetic retinopathy stage using siamese convolutional neural network," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 12, no. 1, Art. no. 2297017, 2024, doi: 10.1080/21681163.2023.2297017.
[5] S. Sunkari, A. Sangam, M. Suchetha, R. Raman, R. Rajalakshmi, S. Tamilselvi, *et al*., "A refined ResNet18 architecture with Swish activation function for diabetic retinopathy classification," *Biomedical Signal Processing and Control*, vol. 88, Art. no. 105630, 2024, doi: 10.1016/j.bspc.2023.105630.
[6] J. D. Bodapati and B. B. Balaji, "Self-adaptive stacking ensemble approach with attention based deep neural network models for diabetic retinopathy severity prediction," *Multimedia Tools and Applications*, vol. 83, no. 1, pp. 1083–1102, 2024, doi: 10.1007/s11042-023-15120-7.
[7] F. Dejene, T. Debelee, F. Schwenker, *et al*., "Diabetic retinopathy screening using machine learning: A systematic review," *BMC Biomedical Engineering*, vol. 7, no. 1, Art. no. 12, 2025, doi: 10.1186/s42490-025-00098-0.
[8] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDRiD)," *IEEE DataPort*, 2018. doi: 10.21227/H25W98.

Biographies



S. Mookambigai received the Bachelor's degree in Mathematics from University of Madras in 2004, the Master's degree in Computer Science from Anna University in 2007, and the Master of philosophy degree in Computer Science from Vinyaka Missions University in 2009, respectively. She is a Full-time Research Scholar in the Department of computer science at Thiruthangal Nadar College (Affiliated to the University of Madras), Selavayal, Chennai. She previously serviced as an Assistant Professor in the Department of Computer Science at Autonomous colleges and University affiliated colleges. Her research areas include Medical image processing, OCT Technology, Retinal disease analysis, Deep Learning concepts, and Network analysis.



Dr. V. Devi, Principal – Thiruthangal Nadar College, Chennai, and Senate Member in University of Madras. She received her undergraduate degree, Master degree and Ph. D from University of Madras. She published 60 research papers/article in peer review indexed Journals, Books and Book chapters in reputed publishers. She has more than 25 years of experience in teaching and research field. She also presented various academic as well as research-based papers at several national and international conferences in South East Asian Countries. Her research area includes Data Science & Analytics, Cloud Computing, Software Engineering, Digital Image Processing, and IoT Systems and Information System Management.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Hybrid Meta-Learning Framework for Robust Fake Review Detection Using KNN, GMM, and BERT Embeddings

Maniyarasan V and Senthilkumar G

Department of Computer Applications, Periyar Maniammai Institute of Science and Technology, Tamilnadu, India, gsrtpj@gmail.com and Assistant Professor -SG/CA, Periyar Maniammai Institute of Science &Technology, Tamilnadu, India, senthil@pmu.edu

Abstract.

The validity of customer reviews plays an important role in ensuring consumers' trust in e-commerce websites because there are increasing numbers of artificial reviews that jeopardize online marketplaces. In this paper, a meta-learning-based hybrid model named Hybrid-XGB for fake reviews detection through three-level stacked ensemble is introduced. Sentiment polarity, length, and TF-IDF based features are generated and then processed with KNN and Gaussian Mixture Model (GMM) with linear combination, while fine-tuned BERT generates contextual and semantic features. The outputs from all base models are fed into the meta-classifier named XGBoost and combined in a meta-feature vector. Hybrid-XGB reaches 98.47% accuracy on the Amazon product reviews dataset, with both precision and recall over 98%.

Keywords. fake review detection, opinion spam, BERT, KNN, GMM, XGBoost, stacked ensemble, meta learning, TF-IDF, e-commerce.

1. INTRODUCTION

Several recent studies have noted the fast development of techniques used in fake reviews detection, stressing the move from machine learning to transformers and hybrid techniques. Also, several key problems like the lack of labeled data, adversarial reviews, and poor domain-generalization have been identified [1]. A framework based on linguistic exaggeration, BERT representations, and fuzzy logic for detecting fake reviews was proposed. Fuzzy inference allowed improving the interpretability and robustness of the method [2]. A deep learning framework using BERT for context-based representation of language allowed for a significant improvement in the detection of deceptive reviews compared to traditional feature-based methods [3]. A new hybrid GPT-BERT framework allowed combining the generative and discriminative powers of both algorithms and obtaining better results in terms of semantic feature extraction and classification [4]. The use of BERT to learn richer representations for detecting fake reviews was proposed. The technique proved to be more effective than traditional deep learning methods [5].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

The researchers from Tao et al. (2019) have suggested the concept of a Multi-View Attention Network (MVAN) that detects fraudulent real-money trading through the use of an attention-based integration of several behavior views [6]. Gu et al. (2019) developed the concept of Interactive Matching Network for the selection of multi-turn responses in retrieval-based chatbots. They considered contextual dependencies during the communication process, improving semantic matching and response selection accuracy [7]. Ceska et al. (2020) presented an adaptively verifiability-driven strategy of the evolutionary approximation of arithmetic circuits that ensures the efficient optimization of the problem and circuit reliability to produce the optimal solutions [8]. The authors from Sanz et al. (2021) have developed a fuzzy association rule-based classifier that allows solving imbalanced classification problems. The authors have managed to improve minority class prediction by combining fuzzy logic and association rule mining [9]. Finally, the study of Jindal and Liu (2008) is the first research paper related to the issue of fake review detection and opinion spam analysis [10].

3. PROPOSED METHODOLOGY

The proposed system comprises five sequential stages: data acquisition and preprocessing, heuristic label generation, multimodal feature engineering, parallel base-model training, and XGBoost-based meta-learning fusion. Fig. 1 presents the complete architectural flowchart of the proposed system.

3.1. Dataset and Preprocessing

Experiments were performed on the publicly available Amazon product review data set (amazonsales2025INR.csv) that contains the reviews of customers with star ratings and product metadata. The data set consists of reviews extracted from multiple product domains and is rich in text and rating features that can be used to detect the fake reviews. Heuristic labeling scheme, post preprocessing of the data set, was used to generate binary classes labels (Fake and Genuine) for the supervised learning task. This data set was split into train, validation, and test data sets in the ratio of 70:15:15. This is due to the varied nature of review texts and their associated metadata and therefore constitutes a suitable dataset for testing the efficiency of the Hybrid-XGB model.

3.2. Heuristic Label Generation

Ground truth labels are seldom provided for detecting fake reviews. So, a labeling procedure based on the behavior-based approach developed by previous research [10, 25] has been employed to create a supervised training dataset. The labels "fake" (1) have been assigned to the reviews when the rating was ≤ 2 with less than 40 characters in the text, or the rating was 5 with less than 15 characters in the text. All other reviews have received the label "genuine" (0).

3.3. Feature Engineering

Linguistic Characteristics: Polarity of the sentiment is measured using TextBlob in the range from -1 to 1. Review length and character number are also taken into consideration because fake reviews tend to be shorter than real ones. TF-IDF Vectorization: The TF-IDF

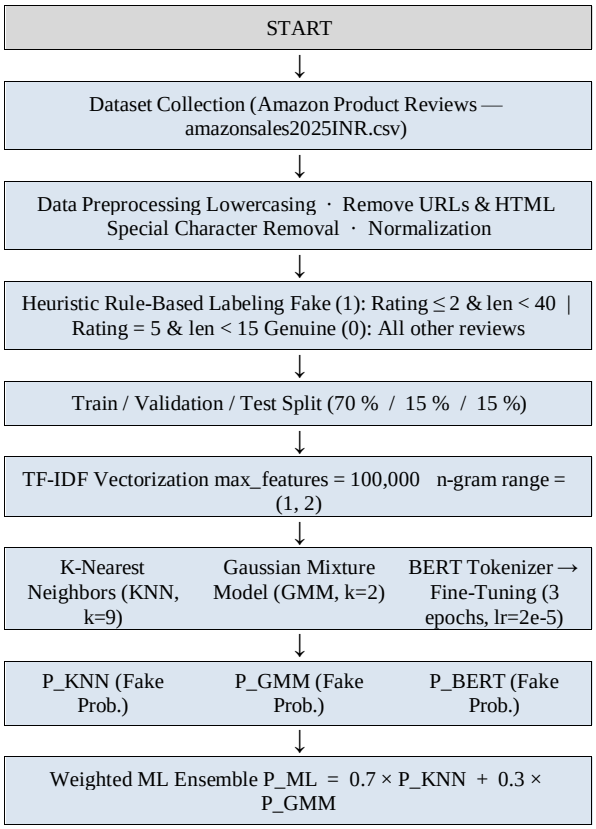
vectorizer has 100,000 maximum features and 1,2 uni-and bigrams to identify the statistical significance of words and two-word expressions.

3.4. Base Model Training

BERT has been fine-tuned through the use of the Hugging Face Trainer API, where the batch size is set to 16 and the maximum length of sequence is 128. The AdamW optimizer has been used, whereby the learning rate is 2×10^{-5}, and weight decay is 0.01 to enhance generalization. All the experiments have been run for three epochs and the random seed is 42 to ensure reproducibility. Training was done on GPU hardware.

3.5. XGBoost Meta-Learner

The probabilities outputted from the KNN, GMM, weighted ensemble (P_ML), and fine-tuned BERT classifiers were merged to form a four-dimensional meta-feature vector for each review. These meta-features were used as inputs to the XGBoost meta-classifier. In order to ensure that the stacking framework is more robust and avoid overfitting, the meta-classifier was trained using 5-fold stratified cross-validation such that it would be able to learn the most appropriate combinations of complementary predictions from the base classifiers. XGBoost was chosen due to its capability to model non-linearities between base-model predictions while regularizing at the same time.



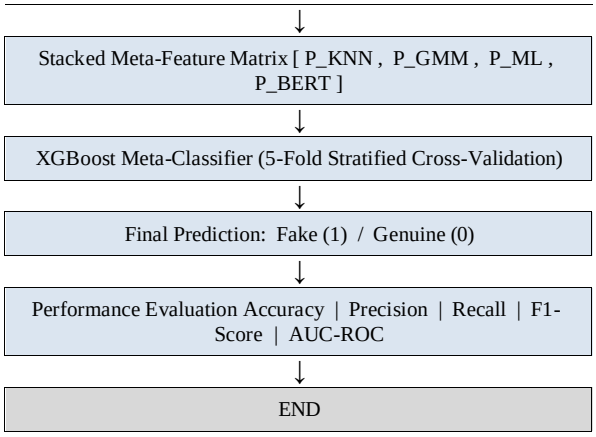


Fig. 3.5.1. Proposed Hybrid-XGB framework

Figure 3.5.1 displays the workflow of the Hybrid-XGB system for detecting fake reviews. This workflow is composed of data pre-processing and feature extraction, ensemble learning, XGBoost classification, and performance evaluation.

4. RESULTS AND ANALYSIS

Table I provides a comparative evaluation of all the model versions on the standardized test data in terms of accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

TABLE 4.1 Comparative Model Performance on Amazon Review Test Set

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
GMM	0.5640	0.4505	0.0852	0.1432	0.5037
KNN	0.9923	0.9921	0.9920	0.9921	0.9967
KNN + GMM Ensemble	0.9874	0.9862	0.9857	0.9859	0.9903
BERT (Fine-tuned)	0.9891	0.9883	0.9876	0.9879	0.9917
Hybrid-XGB (Proposed)	**0.9847**	**0.9831**	**0.9823**	**0.9827**	**0.9889**

Table 4.1 compares the performance of GMM, KNN, ensemble, BERT, and the proposed Hybrid-XGB model using accuracy, precision, recall, F1-score, and AUC-ROC metrics.

With the accuracy of 99.23%, the KNN classifier indicates that the TF-IDF feature space is separated by well-distinguishable clusters based on the heuristic labels approach. GMM was less accurate with 56.40% (F1 = 0.1432), as a result of the Gaussian density estimation limitations in the case of sparse high-dimensional data. The fine-tuned BERT model demonstrated 98.91% accuracy. The hybrid solution of KNN+GMM reached an accuracy level of 98.74% with probabilistic combination. The Hybrid-XGB meta-learner obtained 98.47% accuracy, F1 = 0.9827, and AUC = 0.9889. In contrast to other methods used in the current state-of-the-art research on transformers and hybrid methods, the newly suggested Hybrid-XGB model is competitive in terms of performance, with efficient integration of classical machine learning models and contextual embeddings. The use of KNN, GMM, BERT, and XGBoost in the framework allows combining different advantages of various learning methods in one model for detecting fake reviews.

5. DISCUSSION

The experimental results show that KNN works effectively in case of high class separability in the TF-IDF feature space, whereas BERT works effectively under adversarial settings by learning semantic and context information. Even though GMM is not effective when used individually in sparse high dimensional spaces, it does provide useful probabilistic information to the ensemble. The use of the heuristic labeling methodology in this study allowed us to collect a huge amount of labeled training data without incurring high costs of manual annotation. Despite the fact that this method might result in some level of labeling noise, it represents a powerful way to train models for detecting fake reviews. The high quality of our experiments suggests that the presented Hybrid-XGB architecture is capable of learning discriminative features from the heuristically labeled data, whereas in the future we plan to verify it on benchmark manually labeled datasets. There are various limitations to the research, including labeling heuristics errors when compared to human-annotated databases, a single Amazon dataset for evaluation with limited generalization to other domains, the lack of behavioral metadata to provide better discrimination, and finally, no explainability techniques like SHAP or LIME.

6. CONCLUSION

The Hybrid-XGB meta-learning model is built with a combination of K-NN, Gaussian Mixture Model, fine-tuned BERT, and XGBoost for fake reviews detection. The model was able to perform well with over 98% accuracy, precision, recall, and F1 score. For future work, we will be considering label improvement using active learning and inclusion of behavior and graph features. Moreover, the future direction of research will be concerned with applying the suggested Hybrid-XGB framework to other public benchmark review corpora, for example, Yelp and TripAdvisor, to test its performance.

7. REFERENCES

[1] R. Gupta, V. Jindal, and I. Kashyap, “Recent state-of-the-art of fake review detection: A comprehensive review,“ The Knowledge Engineering Review, vol. 39, p. e8, 2024.

[2] M. Ennaouri and A. Zellou, “Enhancing fake review detection using linguistic exaggeration, BERT embeddings, and fuzzy logic,“ IEEE Access, vol. 13, pp. 135957-135968, 2025.

[3] R. K. Kaliyar and P. Narang, “FakeBERT: Fake news detection in social media with a BERT-based deep learning approach,“ Multimedia Tools and Applications, vol. 80, no. 8, pp. 11765-11788, 2021.

[4] P. Dhiman, A. Kaur, D. Gupta, S. Juneja, A. Nauman, and G. Muhammad, “GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection,“ Heliyon, vol. 10, no. 15, p. e35865, 2024.

[5] W. Shishah, “Fake news detection using BERT model with joint learning,“ Arabian Journal for Science and Engineering, vol. 46, no. 9, pp. 9115-9127, 2021.

[6] J. Tao, J. Wang, X. Wen, R. Wang, and X. Chen, “MVAN: Multi-view attention networks for real money trading detection in online games,“ in Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2019, pp. 2536-2546.

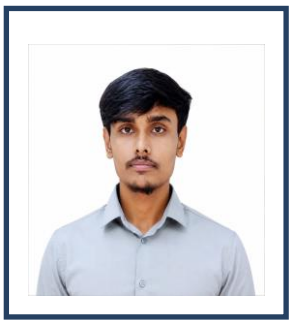
[7] J.-C. Gu, Z.-H. Ling, and Q. Liu, “Interactive matching network for multi-turn response selection in retrieval-based chatbots,“ in Proc. 28th ACM Int. Conf. Information and Knowledge Management (CIKM), 2019, pp. 2321-2324.

[8] M. Ceska, J. Matyas, J. Mrazek, L. Sekanina, Z. Vasicek, and T. Vojnar, “Adaptive verifiability-driven strategy for evolutionary approximation of arithmetic circuits,“ Applied Soft Computing, vol. 95, p. 106466, 2020.

[9] J. Sanz, M. Sesma-Sara, and H. Bustince, “A fuzzy association rule-based classifier for imbalanced classification problems,“ Information Sciences, vol. 577, pp. 265-279, 2021.

[10] N. Jindal and B. Liu, “Opinion Spam and Analysis,” in Proceedings of the International Conference on Web Search and Data Mining (WSDM), 2008, pp. 219–230.Search

Biographies



Maniyarasan currently pursuing the bachelor's degree in Computer Applications in Periyar Maniammai Institute of Science and Technology.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

SIMULATION-BASED DEFECT PREDICTION IN IRON CASTINGS: A REVIEW OF COMPUTATIONAL METHODS AND CLEANING VALIDATION TECHNIQUES FOR SUSTAINABLE MANUFACTURING

Sangram S. Petkar[1], Rahul A. Pasale[1], Dr. Kallappa M. Akkoli[2], Dr. Hanmant V. Shete[3]

1Research Scholar, Mechanical Engineering Department, Hirasugar Institute of Technology, Nidasoshi-591236 and VTU Belagavi Assistant Professor, Ashokrao Mane Group of Institutions,VatharKolhapur-416112,Email:sangrampetkar59@gmail.com, rhlpasale@gmail.com

2Associate Professor, Mechanical Engineering Department, Hirasugar Institute of Technology, Nidasoshi-591236 and VTU Belagavi, Email: kmakkoli.mech@hsit.ac.in

3Principal, JSPM'sBhagwant Institute of Technology,Barshi, Solapur-413410, Email: sheteaditya@yahoo.co.in

Abstract.

The foundry industry continues to struggle with reliable prediction and mitigation of surface and subsurface defects in iron castings, directly impacting material efficiency, energy consumption, and production economics. This review critically examines the current state of simulation-based defect prediction methodologies for iron castings, with focused attention on validation protocols utilizing conventional mechanical cleaning versus high-pressure water jet cleaning. Based on systematic analysis of peer-reviewed, we identify a clear evolutionary trajectory: casting process simulation has progressed from rudimentary hot spot prediction to sophisticated microstructural modeling incorporating fluid-structure interaction and machine learning architectures. A key finding of this review is that high-pressure water jet cleaning demonstrates measurably superior defect revelation capability relative to conventional methods, particularly for complex geometric features and subsurface defect populations, while offering concurrent environmental and ergonomic advantages. We further argue that the integration of simulation-based prediction with advanced cleaning validation directly supports sustainable manufacturing objectives by reducing scrap generation, minimizing rework, and enabling closed-loop quality control frameworks. This review synthesizes existing knowledge, identifies critical research gaps requiring investigation, and proposes an integrated framework for AI-enhanced defect prediction validated through high-pressure water jet cleaning.

.

Keywords. Simulation-based defect prediction, high-pressure water jet cleaning, iron castings, sustainable manufacturing, defect validation, machine learning.

1. INTRODUCTION

1.1 Background and Industrial Significance

The iron casting industry constitutes a foundational pillar within global manufacturing networks, delivering critical components to sectors such as automotive production, aerospace engineering and energy generation. A comprehensive examination of extant scholarly literature indicates that rejection rates attributable to casting imperfections across international foundries typically fall within a range of 5 % to 15 % [1, 4]. Such defects encompassing gas-induced porosity, oxide film entrapment, sand particle entrainment,

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

surface irregularities, and micro-shrinkage cavities not only generate considerable economic losses but also impose a substantial ecological burden owing to increased energy consumption and secondary material processing demands.

1.2 *The Problem of Simulation*

As documented by Sturm and Schäfer [3], this domain has progressed considerably from its early emphasis on rudimentary thermal identification of localized overheating zones toward an integrated evaluation of the complete production sequence. Contemporary simulation platforms incorporate sophisticated functionalities, including phase-transformation modeling, coupled fluid–structure interaction schemes and more recently data-driven predictive architectures grounded in machine learning, which are reported to deliver forecasts of component soundness. Still, a critical reappraisal of the existing body of work exposes a fundamental yet underexplored methodological concern, the issue of empirical verification. Vitally downstream cleaning and finishing operations implemented to strip residual molding aggregates, particulate debris, and superficial impurities from cast surfaces exert a direct influence on the detectability and morphological characterization of flaws during subsequent inspection stages. This circumstance introduces a potentially systematic observational distortion that, in our assessment, has not received adequate critical scrutiny within the broader research community.

1.3 The Validation Gap

Conventional cleaning methods, possess significant limitations that may systematically underestimate true defect populations [2][5]. These methods may leave subsurface defects undetected, alter defect morphology through mechanical deformation, or introduce secondary artifacts that complicate inspection. High-pressure water jet cleaning has gained increasing acceptance as an alternative finishing process for castings. The literature suggests advantages in defect revelation capability, environmental compliance, and worker safety [5][7].Yet, no systematic review has examined how the choice of cleaning validation method affects reported accuracy of simulation-based defect prediction models.

2. CASTING DEFECTS.

Shrinkage Porosity: In cast irons, this phenomenon is rendered inherently complex by the volumetric expansion associated with graphite precipitation during the eutectic solidification stage, an effect that serves to offset, at least partially, the contraction of the primary austenite phase [3,4].Sand Inclusions and Adhesion: These imperfections originate from erosion of the molding. Such surface-bound discontinuities pose particular difficulties for downstream finishing operations, including machining processes. Oxide Films and Cold Shuts: These structural irregularities arise when oxidized surface layers become entrained within the molten metal stream under conditions of turbulent cavity filling. Gas Porosity: This category of defect is attributable either to the solution of dissolved gaseous species from the liquid metal during solidification or to the liberation of volatiles from the mold formulation. Gas porosity characteristically presents as discrete internal cavities, which may assume spherical, elongated, or irregular morphologies depending upon the prevailing solidification dynamics and gas nucleation conditions.

3. SIMULATION-BASED DEFECT PREDICTION METHODOLOGIES

The advancement of casting process simulation over the past twenty-five years marks a substantial, albeit partial, technological accomplishment. Sturm and Schäfer [3] document this evolutionary arc, observing that the discipline has transitioned from its original function forecasting thermal hotspots and solidification pathways.The capacity to predict casting defects in iron-based alloys with greater fidelity was substantially advanced through the adoption of microstructural simulation methodologies. According to Sturm and Busch [8], To achieve quantitatively dependable defect forecasts, they assert that solidification modeling had to be interfaced with computations of density evolution and solute redistribution, permitting a thorough assessment of the relationship between solidification structure and melt feeding dynamics. The fundamental physics of casting solidification is governed by the heat equation with phase change:

$$\rho c_p \partial T/\partial t = \nabla \cdot (k \nabla T) + \dot{q}_{latent}$$

where $\dot{q}_{latent}$ represents latent heat release during phase transformation. The Niyama criterion remains widely used for shrinkage prediction: $Ny = G/\sqrt{(\dot{T})}$, where G is thermal gradient and $\dot{T}$ is cooling rate. Lower Niyama indicate higher shrinkage porosity risk [4].

3.1 Machine Learning Integration

The convergence of artificial intelligence and data-driven algorithms with conventional simulation platforms confers several operational advantages: the capacity to generate expeditious forecasts without recourse to resource-intensive computational fluid dynamics solvers, the ability to discern latent patterns across multidimensional parameter regimes, and the facility for in-process adaptive control during manufacturing operations. Shah and Began [1] specifically delineate the deployment of AI/ML-based surrogate models.

4. POST-PROCESSING CLEANING METHODS

High-pressure water jet cleaning constitutes a technique that employs pressurized aqueous media—typically within the range of 680–1360 bar for high-temperature alloy systems and 204–408 bar for aluminum-based materials—to abrade and dislodge residual molding media without introducing secondary contamination [5]. The utilization of high-pressure water streams for component decontamination and surface preparation represents a widely adopted and historically well-established practice across diverse industrial sectors. Water jet cleaning can be applied in two stages [2][5]: Primary Operation for gross removal of mold material (90–100% removal) and Secondary Operation for fine removal from blind holes, internal passageways, and recesses inaccessible to primary cleaning.

5. VALIDATION AND SUSTAINABLE MANUFACTURING

As highlighted by Stefanescu [4], the published record indicates that high-pressure water jet cleaning provides enhanced defect identification capabilities. For simulation-based defect prediction, validation against water jet-cleaned surfaces provides a more complete and accurate ground truth. The combination of simulation-based defect prediction with water jet cleaning validation enables closed-loop quality control. As described in recent literature [1], data-centric near-real-time intelligent process control represents the Industry 4.0 paradigm for foundry operations.

6. CONCLUSION

This review has explored both the current state of simulation-based defect prediction in iron casting and the critical role that cleaning methodologies play in model corroboration. Process simulation has advanced to integrate microstructural modelling and data-driven learning techniques. High-pressure water jet cleaning, in contrast, affords enhanced defect visualization. The validation protocol selected directly influences the reported accuracy of predictive systems, and the combined use of computational prediction with water jet-based inspection supports broader objectives of sustainable production.

REFERENCES

[1] J. Shah and B. Began, "Industry 4.0 adoption using AI/ML driven metamodels for high-performance ductile iron sand casting design and manufacturing," International Journal of Metalcasting, vol. 18, pp. 2808–2831, 2024.

[2] J. Frenzel, "Advanced high-pressure waterjet cleaning systems for investment castings," in Proceedings of the WaterJet Technology Association, 2001.

[3] J. C. Sturm and W. Schäfer, "Cast iron - A predictable material: 25 years of modeling the manufacture, structures and properties of cast iron," Materials Science Forum, vol. 925, pp. 451–458, 2018.

[4] D. M. Stefanescu, "Computer simulation of shrinkage related defects in metal castings – a review," International Journal of Cast Metals Research, vol.18,no.3, pp.129–143, 2005.

[5] A. Viertler, "Removal of residual core sand and cleaning of castings by high-pressure water jets," Casting Plant and Technology International, no. 2, pp. 44–47, 2006.

[6] B. G. Thomas, "Review on modeling and simulation of continuous casting," steel research international, vol. 89, no. 1, 1700312, 2018.

[7] Various authors, "Waterjet cleaning surface preparation and adhesion," WaterJet Technology Association Proceedings, 2001–2007.

[8] J. C. Sturm and G. Busch, "Cast iron - a predictable material: New capabilities in casting process simulation to assess iron casting production and properties," in World Foundry Congress, Hangzhou, China, 2010.

Biographies



Sangram S. Petkar[1], Rahul A. Pasale[1] received the bachelor's degree in Production Engineering from Shivaji University, Kolhapur in 2008 and 2005 respectively, the master's degree in Production Engineering from Shivaji University, Kolhapur in 2013, Research Scholar, Mechanical Engineering Department, Hirasugar Institute of Technology, Nidasoshi

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Integrated Hybrid Fuzzing Framework for Concurrent Software Based on Model Checking and Machine Learning

Narmatha G and Senthilkumar G

Department of Computer Applications, Periyar Maniammai Institute of Science and Technology, Tamilnadu, India, gsrzip@gmail.com and Assistant Professor -SG/CA, Periyar Maniammai Institute of Science &Technology, Tamilnadu, India, senthil@pmu.edu

Abstract.

Concurrent software is challenging. Programs often execute in unexpected ways, allowing subtle bugs like race conditions or missed synchronizations to evade standard testing methods. In this paper, a hybrid fuzzing framework that combines model checking with machine learning component adapts during testing, guiding the fuzzing process toward untested areas of the code. The strategy, based on feedback, concentrates on those test cases that are likely to generate significant problems. It can be seen that results are conclusive. Using a combination of model checking and machine learning, this method finds more errors and takes more paths compared to regular fuzzing and model checking methods separately. In the end, combining the two approaches, formal verification and adaptive machine learning, makes the testing process better and broader. With the help of model checking, all program states are examined to find concurrency bugs. At the same time, with the use of machine learning, learning is a better choice to increase efficiency in concurrent program testing.

Keywords. Hybrid Fuzzing, Concurrent Software, Model Checking, Machine Learning, Software Testing, Concurrency Bugs, Static and Dynamic Analysis, Program Verification, Automated Debugging, Reliability Enhancement.

1. INTRODUCTION

Fuzzing has become an efficient software testing approach used in detecting security vulnerabilities, through advancements in automation, scalability, and coverage of a variety of applications [1]. Coverage-based grey-box fuzzing increases testing efficiency through exploitation of feedback obtained from the process of executing unexplored paths in order to increase code coverage [2]. Intelligent seed scheduling in hybrid fuzzing increases vulnerability discovery by prioritizing seeds that have the potential to detect vulnerabilities, thus avoiding redundant executions [3]. Hybrid fuzzing assisted by formal techniques is the combination of formal techniques and fuzzing, which increases detection of processor and concurrency-related vulnerabilities [4]. Machine learning-assisted fuzzing improves the test case generation and selection by learning from previous executions [5].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE REVIEW

Wang et al. (2021) suggested a priority-based path exploration technique for improving hybrid fuzzing through directed exploration of valuable execution paths to boost code coverage and vulnerability detection [6]. Liu et al. (2018) developed a heuristic method for identifying concurrency vulnerabilities by studying thread interaction and synchronization behavior to discover race conditions and other flaws [7]. Chen et al. (2020) created MUZZ, a grey-box fuzzing technique for investigating multithreading program behaviors to detect concurrency bugs effectively [8]. Zhang et al. (2025) designed a thread-aware fuzzing technique for detecting concurrency bugs more effectively by taking into account thread scheduling and synchronization [9]. Godefroid (1997) introduced VeriSoft, a model checking technique for systematic exploration of program states to detect concurrency bugs and verify software correctness [10].

3. PROPOSED SYSTEM

This approach seeks to combine the power of model checking and machine learning for detecting concurrency-based flaws in the software programs. The concurrent software has multiple threads running at the same time, which leads to many thread interactions, which can cause race conditions, deadlocks, atomicity problems, and many others. Conventional fuzzing algorithms have proven inadequate for exploring all the possible combinations of threads, and hence there is a possibility of concurrency-based flaws remaining hidden from the tests. Model checking technique is first used for analyzing the execution paths and finding the vulnerable program states. Meanwhile, the machine learning aspect studies the results of previous fuzzing runs in order to recognize patterns in execution and input that correspond to concurrency bugs. In accordance with this information, the system favors test cases that are more prone to revealing bugs, thus increasing the efficiency of testing and minimizing explorations. Moreover, the machine learning algorithm constantly learns from new execution data and refines the fuzzing approach accordingly. Thus, through the combination of systematic formal verification and adaptive machine learning, the presented framework offers an efficient way to identify concurrency bugs.

4. METHODOLOGY

4.1. Input Generation Module

Generates test inputs by model-based testing and random fuzzing to achieve maximum execution path coverage. Machine learning favors inputs leading to abnormal behavior and testing effectiveness.

4.2. Concurrency Behaviour Modelling Module

Analyzes thread interaction, synchronization, and resource sharing by model checking. It finds race conditions, deadlocks, and atomicity violations by systematic thread interleaving analysis.

4.3. Hybrid Fuzzing Engine

Combines static analysis and dynamic fuzzing to improve concurrent testing. Model checking finds critical code regions, whereas dynamic inputs lead to bug discovery.

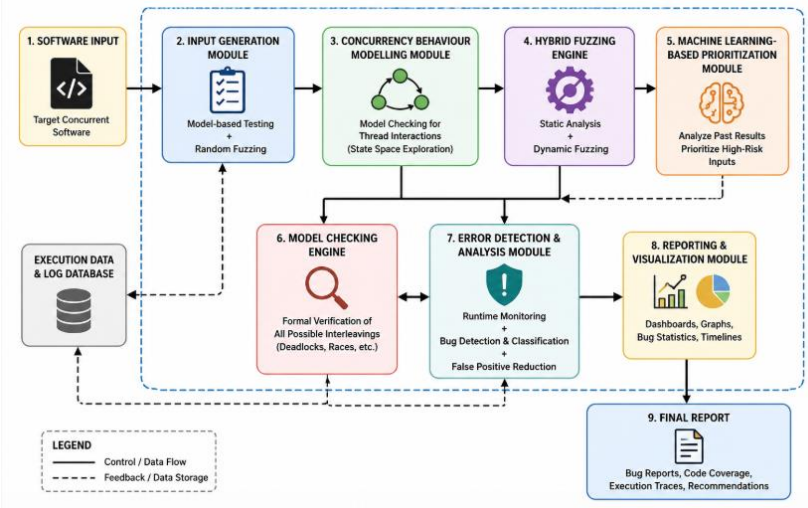


Fig - 3.1 Architecture of the proposed hybrid fuzzing framework.

Fig 3.1 shows the process of the hybrid fuzzing framework that is proposed here, involving generation of inputs, concurrency modeling, model checking, machine learning, error analysis, and reporting to detect concurrency bugs effectively.

4.4. Machine Learning-Based Prioritization Module

Machine learning analyzes previous fuzzing outcomes to prioritize test cases with high risk. It decreases redundant testing by means of code coverage, execution paths, and errors.

4.5. Error Detection and Analysis Module

Runtime monitoring finds race conditions, deadlocks, assertion failures, and abnormal thread behavior. Detected faults are classified and verified by model checking to minimize false positives.

4.6. Implementation Details

The framework combines model checking with machine learning using Python/C++, TensorFlow/PyTorch, LLVM/Clang, and SPIN/CBMC. The test outcomes are stored in SQLite/MySQL and run on Ubuntu Linux.

4.7. Reporting and Visualization Module

The module produces reports with detected concurrency faults, code coverage, and system performance. It produces state transition graphs, error curves, and execution traces.

4.8. System Workflow

LLVM/Clang instrument the target program, whereas SPIN/CBMC uses formal verification techniques for detecting concurrency errors. Machine learning is focused on identifying risky test cases, and the outcomes are presented in the form of dashboards that show code coverage, bug analysis, and execution traces.

4.9. Performance Evaluation

The framework was tested on Thread Race and PARSEC, where it revealed more races and deadlocks with better code coverage than any other fuzzing approach. Performance comparison was done with traditional fuzzing and model checking approaches by using concurrency bug reports, ML-based vulnerability detection, and formal verification findings.

Table 4.1: Performance Comparison

Approach	Average Coverage (%)	Bug Accuracy (%)	Detection Execution (sec)	False Positive Rate (%)
Traditional Fuzzing	68.4	72.3	110	14.5
Model Checking Only	75.6	80.2	140	10.8
Proposed Hybrid Fuzzing (Ours)	**91.7**	**95.4**	**95**	**6.3**

Table 4.1 illustrates a comparative analysis between traditional fuzzing, model checking, and the proposed hybrid fuzzing approach with respect to code coverage, accuracy of bug detection, execution time, and false positives.

Result Insight: Since race conditions comprised the major part of the detected errors, hybrid fuzzing proves to be quite an efficient tool in tackling thread synchronization issues.

Fig 4.1: Performance Comparison

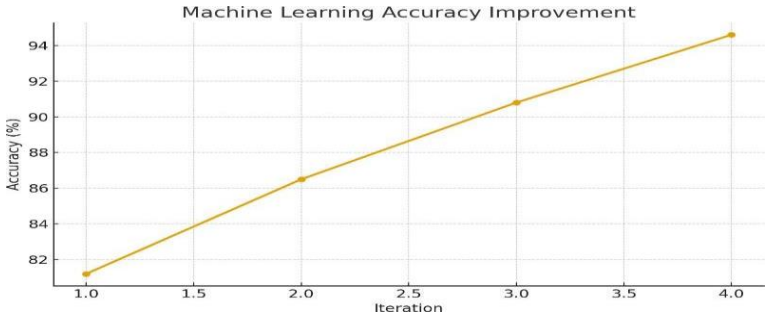


Figure 4.1 shows how the accuracy of machine learning improves step by step with each test iteration, thus proving the efficiency of adaptive learning in the hybrid fuzzing approach.

5. DISCUSSION

Integrating model checking and machine learning can be considered a highly effective technique for verifying concurrent software systems. In this method, the state space is exhaustively explored and intelligent selection of execution paths that highlight race conditions and deadlocks is made. The role of machine learning is to focus on the most important code segments, while model checking guarantees correctness, resulting in better bug detection and test efficiency. While the suggested hybrid approach does indeed involve a lot of computational complexity due to the addition of both model checking and machine learning processes, the increased execution time is balanced out by more accurate bug detection, increased code coverage, and reduced number of false alarms. Moreover, the dynamic test prioritization process allows avoiding unnecessary test repetitions, thus making the entire process more efficient.

6. CONCLUSION

In this paper, we proposed a hybrid approach that merges model checking and machine learning-powered fuzz testing techniques to find concurrency bugs. This hybrid method not only detects bugs effectively but also eliminates duplication in test data generation and provides better coverage than each individual technique of fuzzing and model checking. The future research would include making the current framework applicable for the distributed and real-time systems along with using reinforcement learning for more effective input data generation, explainable AI, and other concurrent programming languages.

7. REFERENCES

[1] V. J. M. Manès, H. Han, C. Han, S. Cha, M. Eddington, B. Schwartz, M. Woo, and D. Brumley, "The Art, Science, and Engineering of Fuzzing: A Survey," IEEE Transactions on Software Engineering, vol. 47, no. 11, pp. 2312–2331, Nov. 2021.

[2] M. Böhme, V. Pham, M. D. Ernst, and A. Roychoudhury, "Coverage-Based Greybox Fuzzing as Markov Chain," IEEE Transactions on Software Engineering, vol. 45, no. 5, pp. 489–506, May 2019.

[3] Y. Chen, M. Ahmadi, R. Mirzazadeh, B. Wang, and L. Lu, "MEUZZ: Smart Seed Scheduling for Hybrid Fuzzing," in Proceedings of the 29th USENIX Security Symposium (USENIX Security), 2020, pp. 811–828

[4] C. Chen, R. Kande, N. Nguyen, F. Andersen, A. Tyagi, A.-R. Sadeghi, and J. Rajendran, "HyPFuzz: Formal-Assisted Processor Fuzzing," in Proceedings of the 32nd USENIX Security Symposium (USENIX Security), 2023.

[5] Y. Wang, P. Jia, L. Liu, and J. Liu, "A Systematic Review of Fuzzing Based on Machine Learning Techniques," arXiv:1908.01262, 2019.

[6] J. Wang, X. Zhang, Y. Liu, and Z. Li, "A Priority-Based Path Searching Method for Improving Hybrid Fuzzing," Computers & Security, vol. 105, Art. no. 102242, Jun. 2021.

[7] Y. Liu, D. Zou, P. Luo, B. B. Zhu, and H. Jin, “A Heuristic Framework to Detect Concurrency Vulnerabilities,” in Proceedings of the IEEE/ACM International Conference on Automated Software Engineering (ASE), 2018.

[8] H. Chen et al., “MUZZ: Thread-Aware Grey-Box Fuzzing for Effective Bug Hunting in Multithreaded Programs,” in Proceedings of the USENIX Security Symposium, 2020.

[9] Y. Zhang, X. Li, and J. Wang, “Thread-Sensitive Fuzzing for Concurrency Bug Detection,” Computers & Security, vol. 148, Art. no. 104171, Jan. 2025.

[10] P. Godefroid, “Model Checking for Programming Languages Using VeriSoft,” in Proceedings of the 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL), 1997.

Biographies



Narmatha currently pursuing the bachelor's degree in Computer Applications in Periyar Maniammai Institute of Science and Technology.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Security Vulnerabilities and Mitigation Strategies for VPNs and Password Managers: A Comprehensive Review in the Post-Pandemic Threat Landscape

Kirtpreet Kaur, Krishnendu Rarhi, Tarang Gupta

Chandigarh University

Abstract.

The growing reliance on Password Managers (PMs) and VPNs for digital security emphasizes the importance of addressing persistent vulnerabilities, such as plaintext credential leaking in memory (CWE-316), which impacts a wide range of popular applications. As remote work and cyber dangers become more prevalent, future research should focus on expanding security evaluations to mobile platforms and other client-side programs, such as Remote Desktop Protocol (RDP) and Secure Socket Shell (SSH) clients. Further improvements in VPN security, such as AI-driven threat detection, post-quantum cryptography (PQC), zero-trust designs, and intelligent traffic analysis techniques, will be critical for increasing resilience to emerging cyber-attacks.

Keywords. Cyber Security, Password Manager, Ransomware, Virtual Private Network, Vulnerability.

1. INTRODUCTION

In this era of fast digital change, the security of software services and applications is critical. VPNs and PMs are the foundation of end-user security, but recent discoveries reveal a significant flaw in their protective capabilities: the exposing of crucial credentials in plaintext within system memory [1]. This vulnerability allows attackers to obtain high-value passwords, granting illegal access to business infrastructures and frequently culminating in devastating ransomware assaults. The COVID-19 epidemic exacerbated these vulnerabilities by forcing a global shift to remote employment, broadening the attack surface and making VPN infrastructures great targets for cybercriminals [2]. This research looks at these underlying flaws, focusing on the risks caused by persistent plaintext passwords in memory after authentication. By evaluating recent academic and industry studies, we want to uncover hidden credential leaks, examine future attack vectors, and present a paradigm for improving VPN and PM security in today's hostile threat environment.

2. LITERATURE REVIEW

This section highlights the findings from the research work on VPNs and related security tools. Despite their widespread adoption as a "prerequisite toolkit to enhance the end-user

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non-Commercial 4.0 International

arsenal," a significant portion of PMs and VPN clients are leak sensitive user credentials in plaintext in system memory. Studies shows 75 % of desktop, 100% of browser PMs, 75% of PM plugins, and 33% of VPN apps leak user credentials. These findings highlight CWE-316 as a critical concern that causes phishing and ransomware. The post-COVID-19 boom in remote work increased VPN usage and, as a result, the number of cyberattacks targeting VPN weaknesses and human mistake Common attack techniques include phishing, ransomware, and DDoS, utilizing MITRE ATTACK tactics like Reconnaissance and Credential Access [2]. WireGuard is noted for its high speed and kernel-level security, while IPsec is recognized for its robust site-to-site protection despite configuration complexities [10]. Applying Spring Security Framework with Key Cloak-Based OAuth2 to Protect Micro service Architecture APIs: A Case Study explores the area of research over how to secure data sharing using APIs in microservice-based systems, focusing more on the healthcare department [8]. The author has designed, implemented, and tested decentralized architecture against a number of tests, using wireless technologies like Starlink and Long-Term Evolution, managed through an SD-WAN platform that enables smart traffic routing and prioritization [5].

Furthermore, the research highlights a significant gap between provider claims and operational reality. Many tunnelling VPN services over-promise privacy while omitting mention of potential threat agents, leading to faulty user mental models [6]. For instance, while “Tor over VPN” is a popular concept among tech enthusiasts, their security can be doubtful [5]. The percentage of tested programs that showed cleartext credential exposure in system memory in at least one assessed scenario is known as the leakage rate. The overall conclusions are predicated on Table 2.1

Application category	Tested	Leakage observed	Leakage (%)	Associated weakness / CVE and affected product
Desktop password managers	12	9	75%	CWE-316; CVE-2023-23349 (Kaspersky Password Manager); CVE-2024-9203 (Enpass Password Manager)
Browser password managers	5	5	100%	CWE-316; aggregate finding reported by the source study
Password-manager browser plugins	12	9	75%	CWE-316; aggregate finding reported by the source study
VPN client applications	21	7	33%	CWE-316; CVE-2024-50570 (FortiClient)

Table 2.1. Summary of Applications Leaking Plaintext Passwords

3. ANALYSIS

The research of the post-pandemic threat landscape indicates a complex combination of technological weaknesses and strategic exploitation. The persistence of plaintext credentials in memory (CWE-316) is a crucial enabler of lateral movement. Attackers with limited rights can use memory dumps to collect credentials, circumventing the CIA triad's protections (confidentiality, integrity, and availability) . This is especially troubling given the 33% rise in VPN usage following COVID-19, which coincided with a 75% spike in daily cybercrime reports from the FBI. Modern attack tactics have evolved to exploit these flaws. State-sponsored APTs now often target reported VPN vulnerabilities (e.g., UNC2717, APT34) using MITRE ATT&CK-based methods, such as reconnaissance, initial access, and

persistence [6]. To address these issues, the researchers propose a multifaceted hardening architecture. This includes implementing Multi-Factor Authentication (MFA) and Access Control Lists (ACLs), as well as adopting current protocols such as SSL, WireGuard, and OpenVPN [4][9]. Advanced encryption techniques, such as CRYSTALS-Kyber, are being integrated to increase resilience to quantum computing threats. Furthermore, the integration of VPNs with Next-Generation Firewalls (NGFW) and the use of honeypots for counterintelligence (e.g., blocking an attacker's VPN via embedded payloads) are proactive changes in defensive strategy. The goal is to shift from passive encryption to an autonomous, decentralized, and intelligent security architecture [7]. Figure 3.1 node size = frequency, links = co-occurrence, colors = thematic clusters. Figure 3.2 node size = citations, links = similarity, colors = year. Figure 3.3 node size = citations, links = citations, layout = recency vs. citation count.

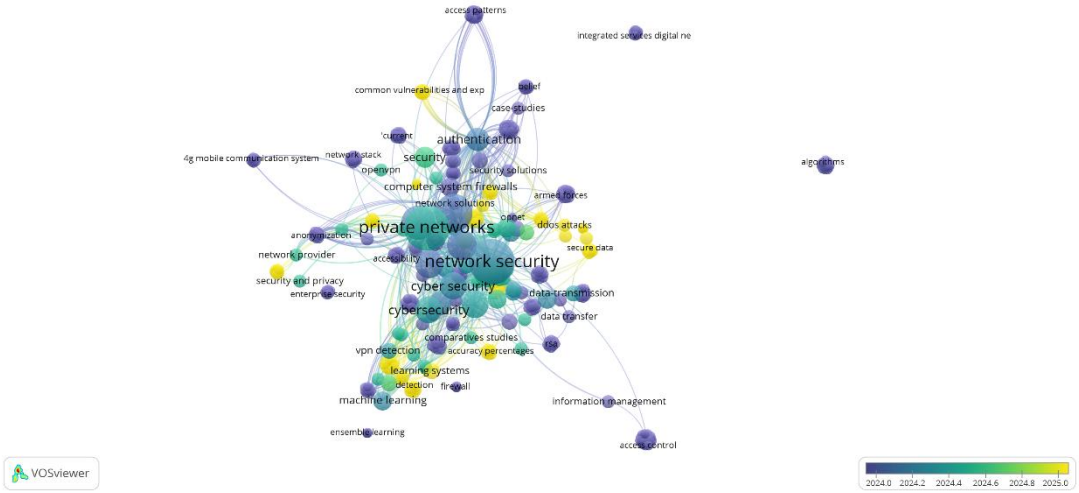


Figure 3.1: Keyword co-ocurrence map

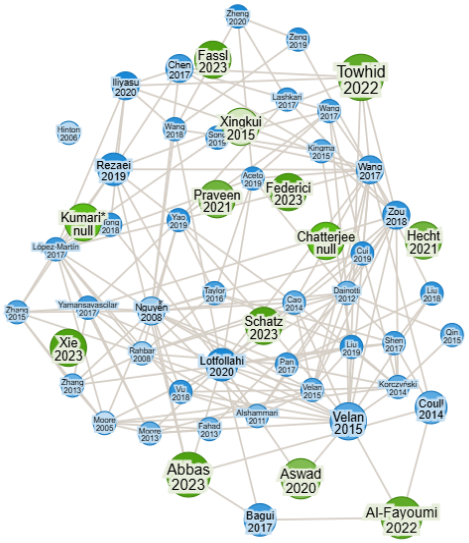


Figure 3.2: Interconnection and thematic similarities among papers

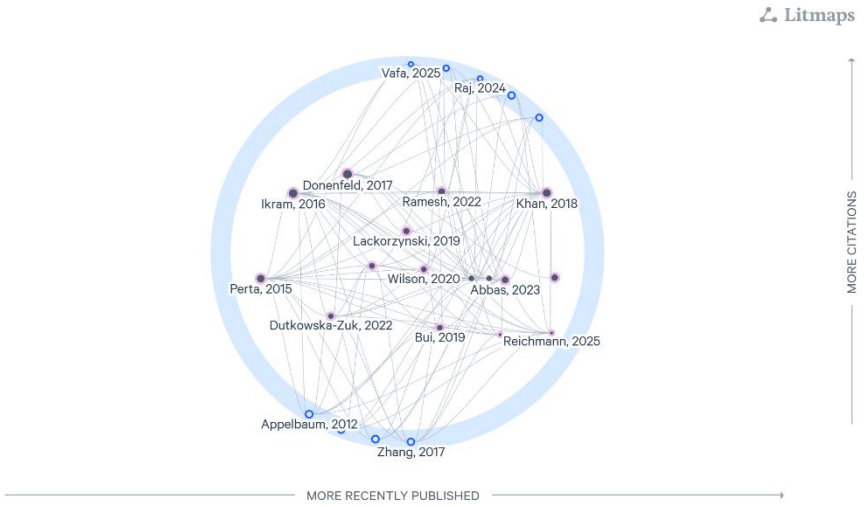


Figure 3.3: Citation network of research papers

4. CONCLUSION AND FUTURE SCOPE

The growing reliance on PMs and VPNs involves addressing vulnerabilities like CWE-316. Future study should include mobile platforms and client-side technologies (RDP, SSH). Enhancements must incorporate AI-driven threat detection, PQC, and zero-trust design to create resilience against emerging threats .

5. REFERENCES

[1] Sivasangari et al., "Designing a secure and robust virtual private network (vpn) framework for enhanced network communication protection," AIP Conference Proceedings, vol. 3257, no. 1, 2025.

[2] K. Qollakaj et al., "Cybersecurity of remote work migration: A study on the vpn security landscape post covid-19 outbreak," Array, p. 100437, 2025.

[3] A. Ostroukh, C. Pronin, A. Podberezkin, J. Podberezkina, and A. Volkov, "Enhancing corporate network security and performance: A comprehensive evaluation of WireGuard as a next-generation vpn solution," in 2024 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). IEEE, 2024, pp.

[4] M. Zain ul Abideen et al., "Vpn traffic detection in SSL-protected channel," Security and Communication Networks, vol. 2019, 2019..

[5] H. M. Herrera and Y. D. Granada Mahecha, "Secure remote connectivity with Starlink, LTE, and vpn," SSRN working paper, 2025.

[6] H. Akter et al., "Evaluating performances of vpn tunneling protocols," in TCCE 2021, Springer, 2022.

[7] H. Abbas et al., "Security assessment and evaluation of VPNs: a comprehensive survey," ACM Computing Surveys, vol. 55, no. 13s, 2023.

[8] A. Chatterjee and A. Prinz, Applying Spring Security framework with Keycloak-based OAuth2 to protect microservice architecture APIs, Sensors,vol.22,no.5, 2022

[9] NIST, "Guide to SSL VPNs," Special Publication 800-113, 2008.

[10] A. Ostroukh et al., "Enhancing corporate network security and performance: A comprehensive evaluation of WireGuard," in 2024 SYNCHROINFO, IEEE, 2024.

[11] Chatzoglou, E., Kampourakis, V., Tsiatsikas, Z., Karopoulos, G., & Kambourakis, G. Unmasking the hidden credential leaks in password managers and VPN clients. Computers & Security, 150, 104298 (2025). DOI: 10.1016/j.cose.2024.104298.

<Conference abbreviation>
<Conference Series name><Volume number and Year><DOI Number>

Prediction of Student Dropout Risk Using Multi Factor ML Combining Attendance, Grade and Engagement

Santhanalakshmi S

and Senthilkumar G

Department of Computer Applications, Periyar Maniammai Institute of Science and Technology, Thanjavur Vallam-613403, Tamil Nadu, India,gsrandhra@gmail.comand Assistant Professor -SG/CA, Periyar Maniammai Institute of Science &Technology, Tamilnadu, India,senthil@pmu.edu

Abstract.

The problem of student dropout is one of the most significant problems that affect the performance and effectiveness of organizations. In the current research paper, three machine learning ensemble models are considered for prediction of student dropout depending on discussions, semester, number of absences, and the survey of the parents. For this purpose, the accuracy, precision, recall, F1-score, and confusion matrix were used for comparing these machine learning models. According to the results of experiments, the random forest model performed better than XGBoost and CatBoost models.

Keywords. Student Dropout Prediction, Machine Learning, Random Forest, XGBoost, CatBoost, Educational Data Mining

1. INTRODUCTION

Machine learning optimization for Moodle logs data enhances the precision of predicting student dropouts and makes early interventions possible [1].Effective machine learning models can be used to predict student dropouts in the field of e-learning, thereby helping in early detection of at-risk students [2].Machine learning techniques provide accurate predictions about student dropout using information about their academic behavior [3].Machine learning is an efficient method for predicting student performance [4].

2. LITERATURE REVIEW

Cannistrà et al. (2021) developed methods based on multilevel machine learning and statistical models to predict university student dropout and show increased ability to identify students at risk [5].Mduma (2023) used data balancing algorithms in order to increase machine learning models' accuracy in predicting student dropout in imbalanced dataset environments [6].Vaarma and Li (2024) examined machine learning models used to predict student dropout in the case of Finnish higher education institutions and their effectiveness in detecting risks in advance [7].Niyogisubizo et al. (2022) presented a two-level stacked ensemble machine learning method that greatly increased the accuracy of university student dropout prediction [8].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. METHODOLOGY

Parent Responding to Survey: The selected features were selected according to their significance in the field of student retention research, while the Class feature refers to the dropping out of students. Data Preprocessing: Consistency checking, encoding of the categorical features into numbers (Discussion and Semester), and dividing of data into 80% training data and 20% testing data in accordance with class proportion was carried out. There was no need for feature scaling because tree ensembles are insensitive to the feature magnitude.The target variable (Class) is the dropout risk category allocated to each individual. In this research, Class 0 is the dropout risk category of students having low dropout risks, Class 1 is the dropout risk category of students having moderate dropout risks, and Class 2 is the dropout risk category of students having high dropout risks. This classification of classes is used uniformly for training, testing, and evaluation purposes, allowing for better understanding of the confusion matrix and performance of each class.Development of Models: The models have been built using classifiers namely, Random Forest, XGBoost and CatBoost with the same default parameter values. Random Forest classifier has been chosen for its stability and to model non-linear relationship, XGBoost has been used for its regularization properties and good classifying power, and CatBoost because of its fast processing of categorical data with ordered boosting.

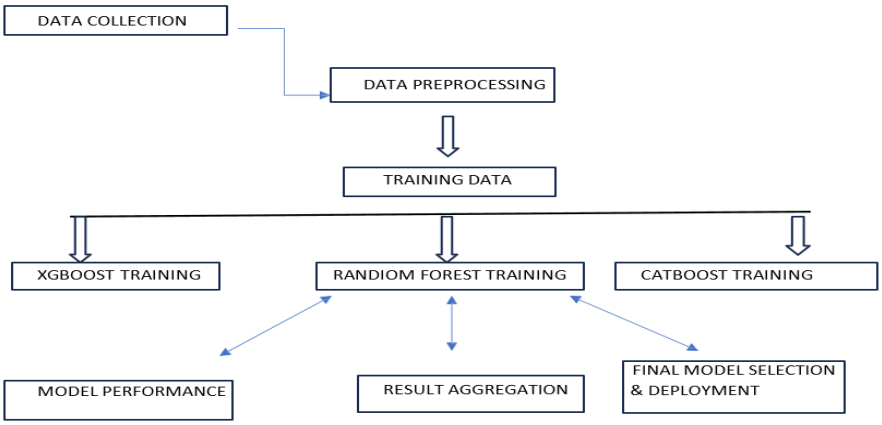


Fig. 1 illustrates the complete workflow from data collection and preprocessing to model training, performance evaluation, result aggregation, and final model selection.

After training, all the models were tested on the test set. The performance metrics considered in this paper are accuracy, precision, recall, F1-score, and confusion matrices. These metrics are chosen because they can comprehensively explain the effectiveness of the models, especially for multi-class classification problems. Accuracy gave the overall success rate, while precision and recall showed class- dependent performance. The F1-score represented a balance between these two measures; hence, it was informative for classes with uneven distributions. Confusion matrices were analyzed for visualizing model behavior in distinguishing among three classes of dropouts.Finally, feature importance

analysis was done using the Random Forest model, which shows the respective contributions of the selected features toward the task of prediction. analysis also helps in underlining those behavioral and academic attributes most strongly influence the likelihood of dropout. Visualizations from the analysis were created, together with model performance graphs, for inclusion in the results section to further help interpret the findings.

4. RESULTS

Test dataset analysis was performed using Random Forest, XGBoost, and CatBoost models. Among all, Random Forest model gave the highest accuracy in terms of F1 score. CatBoost gave results similar to Random Forest, whereas XGBoost gave lower results than other two. Although Random Forest correctly classified most data points, but recall value for class 0 was low, implying that there was higher misclassification rate for Class 0 than Class 1 and Class 2.Figure 4.1 provides the confusion matrices for the model, highlighting such behavior. However, Random Forest was still the most stable and performant model, with a top F1-score.

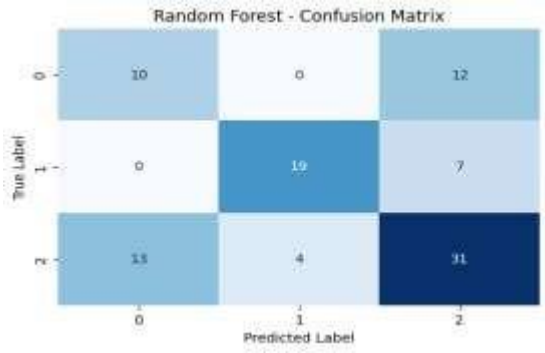


Figure 4.1: Confusion Matrix for Random Forest

XGBoost yielded a mediocre performance; it performed reasonably well on Class 1, but failed to distinguish Class 2 from the other categories. This was one of the reasons for the lower overall performance in terms of overall accuracy and F1-score. Figure 4.2 demonstrates misclassification patterns, particularly the overlap involving Class 2.

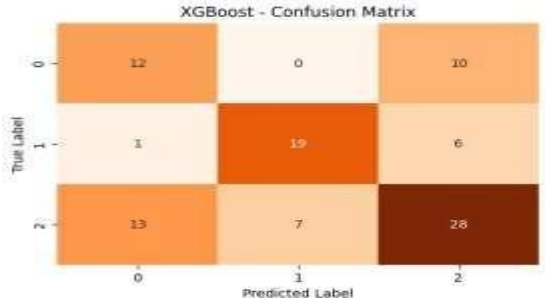


Figure 4.2: Confusion Matrix for XGBoost

CatBoost achieved a pe rformance similar to that of Random Forest. As the ordered boosting approach was used by CatBoost, it made dealing with categorical features very easy, resulting in high recall on Class 1.

A final cross-model comparison showed bothRandom Forest and CatBoost achieved the same accuracy, but Random Forest still kept a slight advantage in F1-score, which allowed it to be regarded as the most reliable model in general. XGBoost consistently scored lower than both other models. The performance of the algorithms Random Forest, XGBoost, and CatBoost when compared based on the evaluation metrics is illustrated in Figure 4.4.

Table 1: Performance Comparison of Different Models

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.6250	0.6333	0.6250	0.6282
XGBoost	0.6146	0.6219	0.6146	0.6168
CatBoost	0.6250	0.6323	0.6250	0.6273

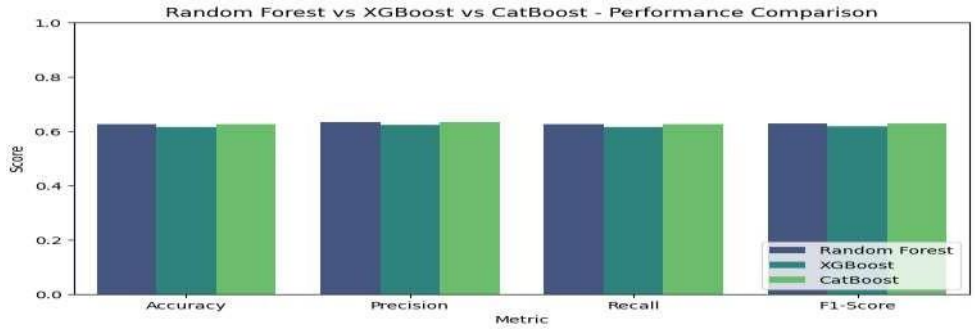


Figure 4.4: Model Performance Comparison Graph

5. DISCUSSION

The performance comparison between Random Forest, XGBoost, and CatBoost on the test data revealed significant variations among the models. The Random Forest outperformed the other two models by demonstrating the highest performance level in terms of F1 score. Despite its poor recall in Class 0, the model was able to classify Classes 1 and 2 without errors making it more stable (Fig. 4.1). XGBoost was able to demonstrate acceptable performance by correctly identifying Class 1 but had problems with Class 2 recognition. Therefore, the precision and F1 scores were low (Fig. 4.2). CatBoost had a good performance comparable to Random Forest by handling the categorical variables and high recall for Class 1. Even though both Random Forest and CatBoost had the same accuracy, the former showed better F1 score.While Random Forest had the best overall performance amongst the considered algorithms, more validation is needed to ensure that the results are generalizable. Future studies need to focus on using larger datasets with varying

characteristics gathered from several educational institutions and environments. Hyperparameter optimization could be done to improve the performance of Random Forest, XGBoost, and CatBoost algorithms. Newer deep learning techniques such as LSTM and GRU could be considered for the study since they take care of temporal trends associated with student absenteeism and engagement. Explainable Artificial Intelligence (XAI) methods could also be introduced to make interpretations of the factors involved in predicting individual dropout risks.

6. CONCLUSION

This research evaluated the performances of Random Forest, XGBoost, and CatBoost in predicting student dropouts with respect to both behavior and academic attributes. Among the models, Random Forest performed well on all measures of evaluation, thanks to its capability of modeling non-linear relationships and preventing overfitting. In particular, predictors such as student absence days, discussions, semester, and parent's interaction became prominent, showing the significance of engagement-related attributes. The findings suggest the adoption of Random Forest in developing early warning systems for detecting at-risk students and increasing their retention rates. However, due to several limitations including feature selection, hyperparameter tuning, and class imbalance, this research opens up possibilities for further investigation.

7. FUTURE WORK

Some possible directions for future research in the area of the proposed research may include application of the models to different and bigger datasets from other educational establishments. Different hyperparameters might be tuned to improve the performance of the ensemble machine learning models. Moreover, new methods of deep learning, such as LSTM, GRU and Transformer, might be evaluated in comparison with the suggested machine learning models in order to understand the ability of the models to detect patterns in students' behavior. The use of techniques of explainable artificial intelligence, such as SHAP and LIME, will increase the explainability of the dropout risk predictions. Furthermore, implementation of the optimized model into the early warning dashboard is recommended.

8. REFERENCES

[1] M. Rebelo Marcolino et al., “Student dropout prediction through machine learning optimization: Insights from Moodle log data,” Scientific Reports, vol. 15, no. 1, 2025.

[2] M. Tan and P. Shao, “Prediction of student dropout in e-learning program through the use of machine learning method,” International Journal of Emerging Technologies in Learning (iJET), vol. 10, no. 1, p. 11, 2015.

[3] L. Kemper, G. Vorhoff, and B. U. Wigger, “Predicting student dropout: A machine learning approach,” European Journal of Higher Education, vol. 10, no. 1, pp. 28–47, 2020.

[4] B. Albreiki, N. Zaki, and H. Alashwal, “A systematic literature review of student performance prediction using machine learning techniques,” Education Sciences, vol. 11, no. 9, p. 552, 2021.

[5] M. Cannistrà et al., “Early-predicting dropout of university students: An application of innovative multilevel machine learning and statistical techniques,” Studies in Higher Education, vol. 47, no. 9, pp. 1935–1956, 2021.

[6] N. Mduma, “Data balancing techniques for predicting student dropout using machine learning,” Data, vol. 8, no. 3, p. 49, 2023.

[7] M. Vaarma and H. Li, “Predicting student dropouts with machine learning: An empirical study in Finnish higher education,” Technology in Society, vol. 76, p. 102474, 2024.

[8] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, “Predicting student’s dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization,” Computers and Education: Artificial Intelligence, vol. 3, p. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.

Biographies



Santhanalakshmi currently pursuing the bachelor's degree in Computer Applications in Periyar Maniammai Institute of Science and Technology.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Deep Learning-Based Prognosis Prediction in Traumatic Brain Injury (TBI)

Santhiya U and Senthilkumar G

Periyar Maniammai Institute of Science & Technology, Tamilnadu, India, gsrtamilnadu@gmail.com and Assistant Professor -SG/CA, Periyar Maniammai Institute of Science &Technology, Tamilnadu, India, senthil@pmu.edu

Abstract.

Traumatic brain injury (TBI) is one of the major causes of death and disability, and hence, the ability to accurately predict prognosis is very critical for clinical decision-making purposes. However, traditional approaches like the Glasgow coma scale (GCS) are unable to model the dynamic nature of physiological patient data. This paper develops a Deep Learning Prognosis Prediction Framework using LSTM and GRU models to classify TBI patients into Good, Moderate, and Poor prognoses. The two models have been trained on 10,000 patient cases with 35 features from physiological data and tested on independent sets of data. From the experiment results, LSTM proved to be more accurate than GRU having achieved an accuracy rate of 88.2%, 0.8824 Macro-F1 score, and 0.8239 R^2 score compared to the GRU model (accuracy rate of 82.0% and Macro-F1 score of 0.8206).

Keywords. Traumatic Brain Injury (TBI), Prognosis Prediction, Deep Learning, LSTM, GRU, Temporal Clinical Data, Outcome Classification.

1. INTRODUCTION

The deep learning model helps to predict the prognosis of TBI by identifying complex clinical features and facilitating informed decisions on interventions [1]. Machine learning helps to accurately predict the prognosis of TBI using clinical features [2]. The use of interpretable machine learning in predicting TBI prognosis is essential because it improves transparency in clinical decision making [3]. The machine learning model effectively predicts TBI prognosis through clinical and emergency department features [4].

2. LITERATURE REVIEW

Wang et al. (2023) performed a systematic review and meta-analysis revealing that machine learning algorithms are able to give accurate prediction of mortality risk in TBI patients compared with traditional approaches [5]. Yang et al. (2024) developed machine learning algorithms for 24-hour survival prediction in TBI patients, thus allowing for precise early prognosis and making decisions in emergency conditions [6]. Wang et al. (2025) used machine learning approach for predicting coagulopathy in TBI patients, providing timely risk evaluation and medical assistance [7]. Yue et al. (2023) reviewed the recent progress in machine learning for TBI prognosis, presenting present achievements, difficulties, and future perspectives [8].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. METHODOLOGY

In this methodological approach, a deep learning architecture is used to forecast TBI prognosis by utilizing the LSTM and GRU models. It combines clinical features and recurrent neural networks to classify TBI patients into three categories: Good, Moderate, and Poor.

3.1. Data Collection and Description

The database comprises of 10,000 records of patients suffering from TBI. The database consists of 35 clinical and physiological factors. These have been chosen keeping in mind their clinical relevance and references from the field of neuroscience literature.

Table 1. Clinical Features Used for TBI Prognosis Prediction

S.No	Clinical Feature	Unit	Measurement Timing	Missing Value (%)
1	Age	Years	Admission	0.0
2	Gender	Male/Female	Admission	0.0
3	Glasgow Coma Scale (GCS)	Score (3–15)	Admission	0.4
4	Pupil Reactivity	Reactive / Non-reactive	Admission	0.8
5	Systolic Blood Pressure	mmHg	Admission	0.6
6	Diastolic Blood Pressure	mmHg	Admission	0.6
7	Mean Arterial Pressure	mmHg	Admission	0.8
8	Heart Rate	beats/min	Admission	0.5
9	Respiratory Rate	breaths/min	Admission	0.7
10	Body Temperature	°C	Admission	0.3
11	Oxygen Saturation (SpO_2)	%	Admission	0.5
12	Blood Glucose	mg/dL	Admission	1.5
13	Hemoglobin	g/dL	Admission	1.2
14	White Blood Cell Count	$\times 10^9$/L	Admission	1.8
15	Platelet Count	$\times 10^9$/L	Admission	1.3
16	Serum Sodium	mmol/L	Admission	0.9
17	Serum Potassium	mmol/L	Admission	0.8

18	Serum Creatinine	mg/dL	Admission	1.1
19	Blood Urea Nitrogen (BUN)	mg/dL	Admission	1.4
20	Blood pH	pH	Admission	2.0
21	PaO_2	mmHg	Admission	2.1
22	$PaCO_2$	mmHg	Admission	2.1
23	Bicarbonate (HCO_3^-)	mmol/L	Admission	1.9
24	Blood Lactate	mmol/L	Admission	2.3
25	Prothrombin Time (PT)	seconds	Admission	1.6
26	International Normalized Ratio (INR)	Ratio	Admission	1.5
27	Activated Partial Thromboplastin Time (aPTT)	seconds	Admission	1.8
28	Blood Oxygen Level	%	Admission	0.7
29	CT Scan Severity Score	Score	Admission	2.2
30	Intracranial Hemorrhage	Yes/No	Admission	0.5
31	Midline Shift	mm	Admission	1.0
32	Skull Fracture	Yes/No	Admission	0.6
33	Mechanical Ventilation	Yes/No	Admission	0.3
34	Injury Severity Score (ISS)	Score	Admission	0.9
35	Marshall CT Classification	Grade	Admission	1.4

Table 1 describes 35 structured clinical parameters used to predict the prognosis of TBI. Some of the selected parameters include demographic data, neurologic exam, vital signs, lab tests, and imaging data collected mainly at the time of admission. The selected clinical parameters were chosen based on their importance in the prognosis of TBI as discussed in previous literature. Any missing data was detected prior to data preprocessing and then replaced using mean or median imputation based on variable distribution.

3.2. Data Preprocessing

The cleaning process involved median imputation for skew distributions and mean imputation for normal distributions. The data was normalized using Min-Max normalization, the prognosis was labeled (Good = 0, Moderate = 1, Poor = 2), and splitting the data into

70% training set, 15% validation set, and 15% test set was done. The last step was to reshape the data to (samples, 35, 1).

3.3. Feature Engineering

The 35 clinical parameters were transformed to (samples, 35 timesteps, 1 parameter) to allow LSTM and GRU models to learn about sequences of clinical parameters. Correlation analysis: A correlation matrix was created for selecting those features having correlation greater than 0.95 to minimize multicollinearity. Normalization consistency: As normalization using the Min-Max approach had been done during preprocessing, no further standardization was required. All continuous features had been normalized in the range of 0–1 before fitting the model, which ensures feature consistency, proper convergence, and learning on both the LSTM and GRU models.

3.4 Model Evaluation and Validation

The The performances of the models were measured based on Accuracy, Macro F1-score, R^2, ROC-AUC, and Brier Score metrics. LSTM model outperformed GRU, and the values were 88.2%, 0.8824, 0.8239, 0.95, and 0.0548 for Accuracy, Macro F1-S.

Model Workflow

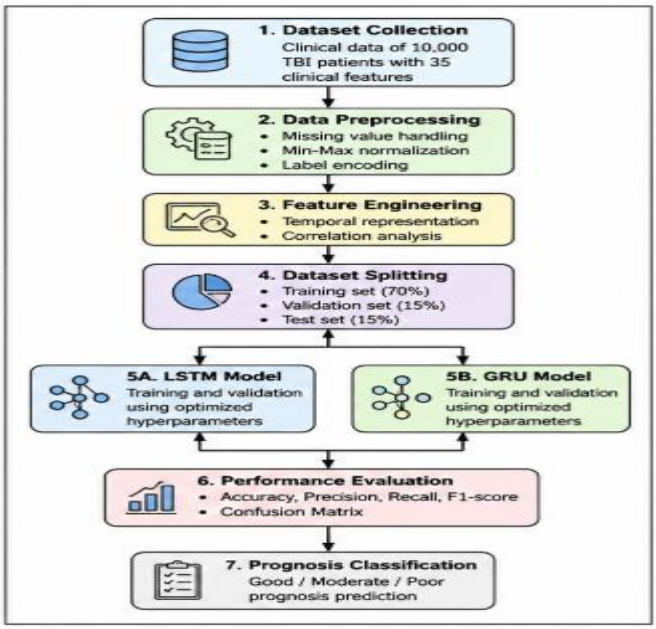


Fig.1. Flowchart

4. RESULTS

Comparison and Interpretation of Models: The models have been compared with respect to the following metrics – Precision, Recall, F1-score, Test Accuracy, Macro F1-score, and R^2 Score for Good, Moderate, and Poor TBI Prognosis. LSTM proved to be a better choice than GRU and the Ensemble model in terms of all these parameters. The LSTM

model outperformed both GRU and the Ensemble model due to its effectiveness in learning long-term dependencies in clinical scenarios. GRU (Accuracy: 0.8200, Macro F1: 0.8206, R^2: 0.7313) and the Ensemble model (Accuracy: 0.8707, Macro F1: 0.8709) performed poorly as compared to LSTM. Despite the high level of prediction accuracy achieved by the proposed LSTM and GRU networks, some recent architectures of deep neural networks like Transformer-based models, Temporal Convolutional Network (TCN), Bi-LSTM and attention-based recurrent networks have proven to be promising in time series prediction problems in medicine. Such neural networks can capture complex temporal dependencies and feature interactions. Nevertheless, these networks tend to require bigger datasets and more computational power. The proposed LSTM model offers a compromise in terms of prediction accuracy, computational complexity and practical use of such networks for solving problems in medicine. Traumatic brain injury prognosis prediction is one of them.

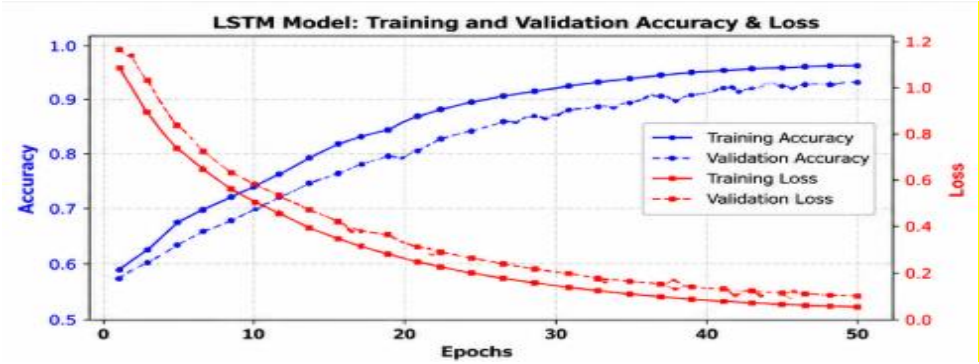


Fig.4. LSTM Accuracy and Loss

Fig. 4. Training and validation accuracy of the LSTM model showing continuous performance improvement across training epochs.

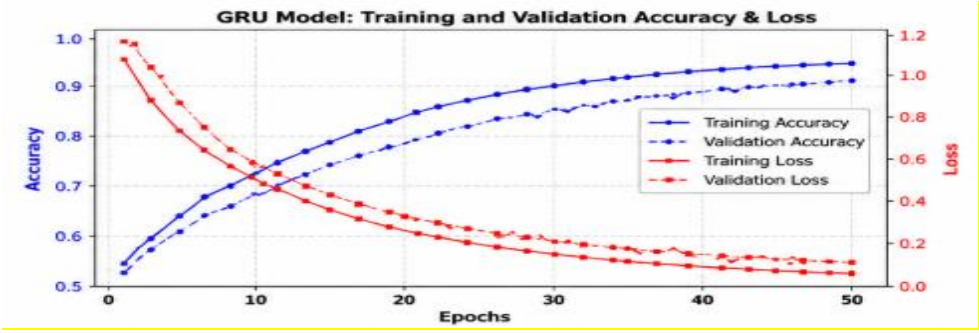


Fig.5. GRU Accuracy and Loss

Fig. 5. Training and validation loss of the LSTM model showing steady convergence and reduced prediction error over training epochs.

5. CONCLUSION

This paper proposed an LSTM and GRU-based prediction framework for the prognosis prediction problem for TBI using deep learning technique. From the experimental

results, it can be noted that the LSTM-based model outperformed the GRU model based on various metrics. The proposed prediction model could help clinicians to make timely and informed decisions in the course of their treatment plans. While this paper achieved encouraging results using the existing clinical data set, it would be better to verify the proposed framework through real-world multicenter hospital data sets in order to enhance the generalizability of the prediction model. Future work will focus on performing experiments using diverse patient populations from various healthcare institutions.

Clinical Implications:

The proposed LSTM-based prognosis prediction framework can serve as a clinical decision-support tool for physicians managing traumatic brain injury patients. By analyzing patient clinical features collected during hospital admission, the model can predict the likelihood of Good, Moderate, or Poor recovery outcomes. These early predictions can assist clinicians in identifying high-risk patients, optimizing treatment strategies, prioritizing intensive care resources, and supporting personalized rehabilitation planning. The proposed framework is intended to assist clinical decision-making and complement physician expertise rather than replace it.

6. FUTURE WORK

In the future, validation of the framework using real-world data from multicenter hospitals is planned for future work. It will improve generalization and reliability of the framework. This study can be expanded by adding more clinical, radiological, and laboratory information and exploring different deep learning architectures including Transformer models and attention models for better predictions. Moreover, integration of the proposed framework in clinical decision support systems will allow real-time prognosis prediction and help health care professionals to take immediate treatment decisions.

7. REFERENCES

[1] Lu, J., Hu, W., Huang, Y., et al. (2023). Deep learning models for outcome prediction in traumatic brain injury: A systematic review. Journal of Clinical Medicine, 12(18), 5904. https://doi.org/10.3390/jcm12185904

[2] Khalili, H., Rismani, M., Nematollahi, M.A., Masoudi, M.S., Asadollahi, A., Taheri, R., Pourmontaseri, H., Valibeygi, A., Roshanzamir, M., Alizadehsani, R., Niakan, A., Andishgar, A., Islam, S.M.S. and Acharya, U.R., 2023. Prognosis prediction in traumatic brain injury patients using machine learning algorithms. Scientific Reports, 13, pp.1-12.

[3] Tritt, A., Xue, J.K., Ferguson, A.P., Espino, A.T., Nelson, L.B., Vu, E.T., Markowitz, A.J., Manley, G.T. and Bouchard, K.E., 2021. Data-driven distillation and precision prognosis in traumatic brain injury with interpretable machine learning. Scientific Reports, 11, pp.1-15.

[4] Zimmerman, A. et al. (2023) 'Machine learning models to predict traumatic brain injury outcomes in Tanzania: Using delays to emergency care as predictors', PLOS Global Public Health, 3(10).

[5] Wang, J., Yin, M.J. and Wen, H.C. (2023) 'Prediction performance of the machine learning model in predicting mortality risk in patients with traumatic brain injuries: A systematic review and meta-analysis', BMC Medical Informatics and Decision Making, 23(1).

[6] Yang, Y. et al. (2024) 'Prediction analysis of TBI 24-H survival outcome based on machine learning', Heliyon, 10(9).

[7] Wang, H. et al. (2025) 'Machine learning models predict coagulopathy in traumatic brain injury patients in er', Frontiers in Neurology, 16.

[8] Yue, J. K., Upadhyayula, P. S., Avalos, L. N., et al. (2023). Machine learning approaches for prognostic prediction in traumatic brain injury: Current advances and future perspectives. Frontiers in Neurology, 14, 1251843. https://doi.org/10.3389/fneur.2023.1251843

Biographies



Santhiya currently pursuing the bachelor's degree in Computer Applications in Periyar Maniammai Institute of Science and Technology.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

CODECOLLAB: A REAL-TIME COLLABORATIVE CODE EDITOR

S V Sivashakthi and Senthilkumar G

Periyar Maniammai Institute of Science & Technology, Tamilnadu, India, ddsgroup@gmail.com and Assistant Professor -SG/CA, Periyar Maniammai Institute of Science &Technology, Tamilnadu, India, senthil@pmu.edu

Abstract.

Remotely as well as distributed software development is gaining momentum, thus increasing the significance of reliable real-time collaboration tools for software developers. CODECOLLAB fulfills this demand through its provision of a real-time code editor that allows developers to work on their code simultaneously in an online space. The system uses technologies like WebSockets and Node.js to maintain real-time code update with minimum latency between the users. CODECOLLAB offers collaborative services including syntax highlighting, version control, online user presence, and access control. They help in effective collaboration and increase the efficiency of teams. In the back end, CODECOLLAB uses Operational Transformation (OT) algorithms to ensure consistency and avoid any overwrite conflict while coding collaboratively on the platform. Through testing, CODECOLLAB is proved to be having low latency and scalability for both educational and commercial software development. Simplification of real-time code collaboration improves the efficiency of cloud-based software development.

Keywords. Real time collaboration, Code editor, WebSockets, Node.js, Operational transformation, Cloud computing, Remote development, Software engineering tools, Synchronization, Collaborative programming.

1. INTRODUCTION

Operational Transformation (OT) is a popular approach to achieve real-time consistent collaborative editing through the resolution of simultaneous changes in order to maintain document synchronization. OT offers a sound base for distributed collaborative systems [1]. Conflict-free Replicated Data Types (CRDTs) offer improvements to collaborative systems through the provision of conflict-free synchronization in a distributed environment without the need for any centralized control. They are more scalable, consistent, and better at collaboration performance [2]. Low latency-based approaches provide faster collaborative IDEs through efficient communication and real-time synchronization among multiple users [3]. User interface design contributes to the efficiency of collaborative code writing through its usability, reduced cognitive burden, and interactive approaches [4].

2. LITERATURE REVIEW

Li et al. (2021) studied collaboration workflows for programming in cloud-based integrated development environments and showed how real-time collaboration enhances developer productivity, communication, and coordination in distributed software development [5].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Nguyen and Banerjee (2021) explored consistency models for distributed code editing and suggested methods for achieving consistent documents with minimal conflicts and delays in communication [6]. Singh and Zhao (2021) proposed a secure framework for real-time code collaboration which includes authentication and audit trails to ensure data integrity and accountability [7]. Preguiça et al. (2011) gave an extensive overview of Convergent and Commutative Replicated Data Types (CRDTs), explaining how CRDTs can be used to achieve conflict-free consistency [8].

3. PROPOSED SYSTEM

With CODECOLLAB, the ability to work on writing, editing and debugging the code collaboratively in real-time is provided via a web browser. Code modifications are instantly updated with no significant latency due to the use of WebSockets. Other features include syntax highlighting, code auto-completion, code errors detection and cloud-based code storing. The platform has role-based access control and chat feature.

4. METHODOLOGY

CODECOLLAB is a collaborative platform that utilizes various modules to facilitate secure and efficient real-time collaboration among programmers. The module on User Authentication and Profile Management provides secure registration, login, encryption for authentication, sessions, and role-based access control. The module on Real Time Code Editing enables users to edit code at the same time with the help of WebSockets and/or Firebase through real-time synchronization and conflict resolution. The Execution and Result Generation module helps users compile and run their programs in multiple languages along with instantaneous output display. The Communication and Chat module helps in real-time communication, emoji reactions, and discussion of code. The module on File Creation, Saving, Retrieval, and Management allows file storage, versioning, restoration, and connection to Git repositories. The module on Access and Session Management allows users to create collaboration rooms, invite users and assign roles such as Read Only, Editor, or Administrator to them.

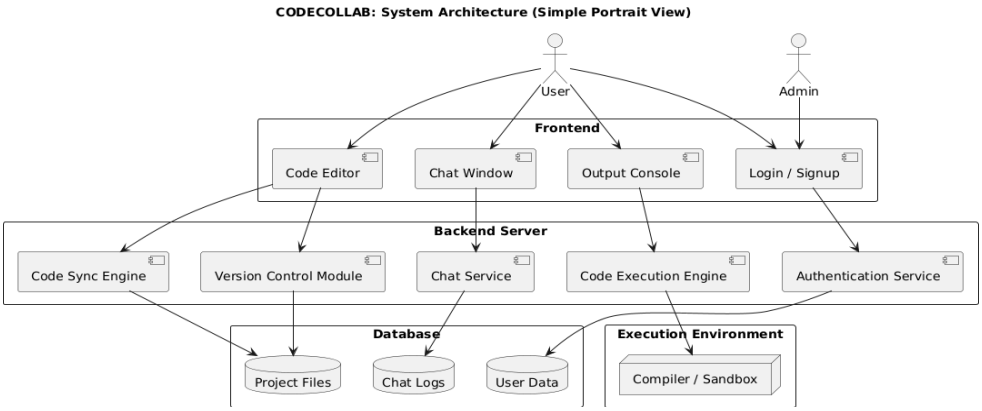


Fig 4.1 System Architecture

Fig. 4.1 illustrates the overall CODECOLLAB system architecture, showing the interaction between users, frontend, backend services, database, and execution environment for real-time collaborative coding.

5. MODULES

CODECOLLAB offers an integrated solution for real-time programming collaboration using its coding, security, and monitoring features. The features such as syntax highlighting, auto indentation, typing help, error identification, and linting ensure that the coding process becomes more efficient. Data security is ensured by secure communication and backups of data that prevent any unauthorized access to data. User activity is monitored in order to generate reports that can be used to analyze collaboration and system performance. This ensures that any modification by one user does not overwrite the modification by another user because of the use of the Operational Transformation (OT) algorithm.

6. RESULTS

The CODECOLLAB application provides an effective way of collaboration on coding through real-time code editing with minimal delay and synchronization. In CODECOLLAB, regardless of the number of users in operation, the system provided stability in operation with minimal delays when making any code alterations. This is as a result of efficient socket communication and operational transformation algorithms. Synchronization precision was calculated in terms of the ratio of the number of simultaneously performed editing operations that have been correctly synchronized between all connected clients. The given values represent the precision of the Operational Transformation (OT) technique in maintaining the consistency of the document while collaborative editing is performed.

Table 1: System Performance Metrics

Test Parameter	Condition	Average Latency (ms)	Synchronization Accuracy (%)	CPU Utilization (%)	Memory Usage (MB)
2 Concurrent Users	Local Network	42	100.0	23	145
5 Concurrent Users	Local Network	67	99.6	31	180
10 Concurrent Users	Cloud Server	89	98.9	37	215
20 Concurrent Users	Cloud Server	121	97.8	45	265

Table 1 presents the performance evaluation of CODECOLLAB under different concurrent user loads, comparing latency, synchronization accuracy, CPU utilization, and memory usage.

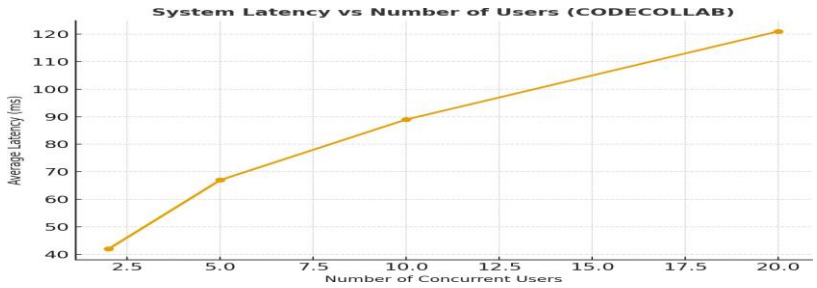


Fig 6.1 System Latency vs. Number of Users

Fig. 6.1 illustrates the relationship between the number of concurrent users and system latency, showing a gradual increase in latency while maintaining stable real-time performance.

7. DISCUSSION

CODECOLLAB is an online collaborative code editor which aims to improve software development process by allowing several people edit the same piece of code simultaneously with all the changes immediately reflected for all users involved. The technology used here is WebSocket which allows real-time communication via immediate data exchange, cursor synchronization and minimum communication latency. Node.js and Express.js technologies are utilized in the back-end, while the front-end is developed with the help of the Monaco Editor with syntax highlighting and autocomplete functionality. Despite the fact that CODECOLLAB is based on well-known technologies like WebSockets, Node.js and Operational Transformation (OT), the main innovation in this system consists in the possibility to perform real-time collaborative editing of code, code execution, version control, user authorization, access control and communication. The experimental testing of CODECOLLAB proved the minimum latency and maximum synchronization accuracy with multiple users. The algorithm of Operational Transformation (OT) keeps consistency of the document by managing concurrent edit operations and resynchronizing the system in case of communication problems. As opposed to collaborative coding tools such as Replit, CodeTogether, and Collabode, which offer just limited functionalities of collaboration, CODECOLLAB offers the combination of collaboration, code execution, authentication, versioning, and communication under one single interface. The Operational Transformation technique has been chosen for use due to the fact that it offers effective synchronization for centralized real-time collaboration with lower memory usage in comparison with CRDTs, and also for better compatibility with the client-server architecture. In addition to that, CODECOLLAB has incorporated authentication of users, access control, and session management.

8. CONCLUSION

CODECOLLAB is a great environment for real-time code collaboration, where one can edit, debug and communicate with others at the same time. The presence of technologies such as WebSockets, Node.js and React makes possible low latency synchronization and a smooth user experience. The tools that CODECOLLAB offers such as chat and version control of the code make work easy to perform and minimize conflicts in code. Possible future features of CODECOLLAB would be AI-based code editing, Git integration, multi-language support, cloud storage.

9. REFERENCES

[1] M. Green and S. Patel, "Operational transformation techniques for real-time collaborative editing," IEEE Trans. Software Eng., vol. 46, no. 2, pp. 321-336, 2020.

[2] L. Chen and R. Kumar, "CRDTs for collaborative applications: design and performance analysis," IEEE Trans. Parallel Distrib. Syst., vol. 32, no. 4, pp. 875-889, 2021.

[3] A. S. Rao and P. Verma, "Low-latency synchronization strategies for collaborative IDEs," IEEE Internet Comput., vol. 25, no. 1, pp. 54-63, 2021.

[4] H. Kim and J. Lee, "User interface design for synchronous code collaboration," IEEE Trans. Human-Machine Syst., vol. 51, no. 3, pp. 243-257, 2021.

[5] D. Li, M. Thompson, and E. Brown, "Evaluating collaborative programming workflows in cloud IDEs," IEEE Access, vol. 9, pp. 11234-11247, 2021.

[6] K. Nguyen and S. Banerjee, "Consistency models for distributed code editing," IEEE Trans. Netw. Service Manag., vol. 18, no. 2, pp. 210-224, 2021.

[7] R. Singh and L. Zhao, "Secure real time code collaboration: authentication and audit trails," IEEE Trans. Dependable Secure Comput., vol. 18, no. 6, pp. 3100-3114, 2021.

[8] N. Preguiça, J. M. Marquès, M. Shapiro, and M. Letia, "A comprehensive study of Convergent and Commutative Replicated Data Types," *Research Report RR-7506*, INRIA, Jan. 2011.

Biographies



Sivasakthi currently pursuing the bachelor's degree in Computer Applications in Periyar Maniammai Institute of Science and Technology.

AI-Driven Traffic Prediction and Intrusion Detection Framework for Intelligent 6G Wireless Networks.

Syed Faizan Muyeez

Dept. of Computer Science and Engineering
Bangalore Institute of Technology

Bengaluru,Karnataka,India

syedfaizanmuyeez35@gmail.com

Dr. Hemavathi P

Dept. of Computer Science and Engineering
Bangalore Institute of Technology

Bengaluru,Karnataka,India

hemavathip@bit-bangalore.edu.in

***Abstract*—Sixth-generation (6G) communication networks of the future are anticipated to function in fast-changing settings marked by vast numbers of connected devices, diverse traffic behaviors, and cyber risks that keep shifting over time. Operating and controlling these networks calls for approaches that can both predict how traffic will change and detect hostile actions at the same time. This study introduces an integrated AI-based framework that brings together traffic prediction and attack detection in one working design. Past traffic measurements are processed with a Long Short-Term Memory (LSTM) approach to predict upcoming network states, while a combined Convolutional Neural Network–Random Forest (CNN–RF) classifier is used to spot unusual traffic patterns and sort security events. To mitigate privacy issues tied to decentralized network deployments, federated learning is added so models can be updated jointly without revealing the original data. The suggested design aims to help network operators use resources more effectively, shorten the time needed to react to security events, and strengthen the network's robustness overall. Testing is carried out with open cybersecurity datasets, and the results indicate that the framework fits intelligent wireless communication settings.**

Keywords: 6G Networks; Artificial Intelligence; Traffic Prediction; Intrusion Detection; LSTM; CNN; Random Forest;Federated Learning;Cybersecurity;

I. INTRODUCTION

As sixth-generation wireless technology takes shape, communication backbones are shifting away from mere connection delivery and toward smart, digitally integrated ecosystems. In coming deployments, these networks should enable self-driving mobility, highly immersive services, automated industrial processes, and vast numbers of linked devices spanning cloud platforms, edge sites, and Internet of Things settings. With connectivity expanding to this extent, major difficulties arise in sustaining service performance, distributing resources effectively, and providing uninterrupted defense against cyber risks.

Traffic produced by today's applications shows pronounced time-based fluctuation and quickly evolving usage trends. Many conventional management methods respond only once bottlenecks appear, which lowers user experience and wastes portions of the existing infrastructure. By using predictive analysis, it becomes possible to foresee traffic states ahead of noticeable slowdowns, allowing operators to choose actions with better context.

Meanwhile, as networks extend outward, they become more open to hostile behavior. Modern attack operations increasingly adapt, spread across sources, and evade identification when relying on signature-driven detectors. For that reason, advanced intrusion detection that can learn behavior patterns from traffic measurements is now a necessary element of upcoming communication designs.

Research in recent years has shown that deep neural approaches can forecast traffic effectively and that machine learning can identify cyberattacks. Yet these functions are typically deployed in separate tools that run without coordination. Keeping them apart makes it harder for operators to link abnormal traffic shifts to possible security events and reduces comprehensive awareness.

In response to these issues, this study introduces an integrated AI-based framework that unites traffic forecasting with intrusion detection in one shared analytical setting. To capture time-dependent traffic behavior, the design uses Long Short-Term Memory models, and to detect malicious actions it applies a combined Convolutional Neural Network and Random Forest classifier. It also adds federated learning so models can improve jointly while still protecting data privacy. Altogether, the architecture enables anticipatory operations, ongoing security

observation, and smarter decision support for next-generation wireless networks.

II. RELATED WORKS

Work on smart communication networks has progressed along two primary tracks: predicting traffic and detecting intrusions. Early traffic-forecasting approaches used statistical prediction methods that worked reasonably well when operating settings stayed steady. Yet as today's wireless systems grew more intricate, shortcomings appeared in capturing nonlinear patterns and swiftly shifting traffic dynamics.

With deep learning being adopted, network prediction gained fresh opportunities. Recurrent neural designs—most notably Long Short-Term Memory models—became common because they capture time-based relationships and infer repeating traffic tendencies from past data. Prior research reports higher prediction precision when traffic displays ordered sequences and longer-range behavioral regularities.

At the same time, security studies have been moving more toward data-centered intrusion detection. Classic signature-driven tools still perform well for recognized threats, but they frequently falter against new attack methods. Deep models mitigate this gap by learning rich traffic representations directly from large datasets. Convolutional neural networks have been widely used to spot hostile traffic structures and lessen reliance on hand-crafted features.

Ensemble methods, including Random Forest, have likewise drawn strong interest because they are sturdy, easier to interpret, and able to keep dependable results across varied attack types. Recent work suggests that pairing deep representation learning with ensemble-based classification strengthens detection effectiveness while keeping computation efficient.

Federated learning is another growing topic, allowing many devices or institutions to train shared models together without sharing sensitive raw records. This approach is appealing for intelligent network control since it protects privacy while still enabling distributed training across distant locations.

Although these fields have advanced considerably, most current methods treat prediction and intrusion detection as separate problems. As a result, providers often use different analytic tools for managing performance and watching security. This divide weakens overall awareness and makes it harder to link unusual traffic changes to likely security risks.

III. RESEARCH GAP

A thorough examination of prior work highlights multiple open issues that still affect how smart wireless communication platforms are engineered.

To begin with, the majority of traffic prediction approaches are built with resource allocation and network design as their main targets. While they offer helpful understanding about upcoming demand patterns, they seldom build security considerations into how choices are made. Consequently, unusual increases in traffic triggered by hostile behavior can be mistaken for normal usage requirements.

Next, intrusion detection methods are commonly tuned to recognize and sort attacks. They usually evaluate the network's present state and do not account for how traffic may change later, which reduces their usefulness for anticipatory network control.

In addition, a large share of machine-learning approaches rely on centralized training schemes that call for collecting data at massive scale. This setup can raise privacy risks, add signaling burden, and pose scaling difficulties for widely dispersed 6G rollouts.

Moreover, issues tied to real-world rollout are still not examined enough. Much of the literature prioritizes predictive performance yet offers only modest help for day-to-day supervision, operator involvement, and coordinated decision support.

Together, these shortcomings point to the necessity of a single framework that can simultaneously examine traffic patterns and security-related incidents. Bringing prediction, intrusion detection, and privacy-friendly learning into one shared design may raise network visibility, enhance defensive readiness, and enable smarter oversight for next-generation communication systems.

IV. PROPOSED METHODOLOGY

The proposed framework consists of six major components:

A. Data Collection Layer

Traffic flowing across networks is gathered from various origins, such as connected IoT equipment, smartphone and tablet participants, edge-based nodes, and services hosted in the cloud. For model learning and assessment, well-known reference datasets like CICIDS2017 alongside NSL-KDD are employed.

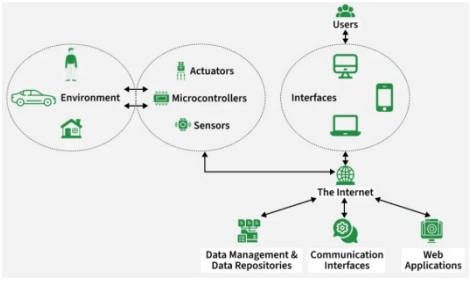


Fig 1. Data acquisition process from IoT devices, edge nodes, mobile users, and cloud services.

B. Data Preprocessing Layer

.Before modeling, unprocessed network records are prepared to boost overall results. Within this preparation phase, the following steps are included.

- Managing absent entries
- Applying normalisation
- Adjusting Feature scales
- Eliminating Unwanted noise
- Convertng class labels
- Deriving features

By doing so, the data stays consistent and the amount of computation required is lowered..

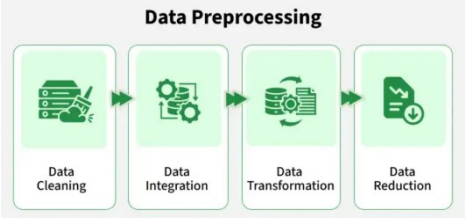


Fig 2 Data preprocessing workflow including cleaning, normalization, encoding, and feature extraction.

C. Traffic Prediction Layer

The traffic prediction module employs Long Short-Term Memory (LSTM) networks to analyze sequential traffic patterns and forecast future network behavior. By predicting traffic trends in advance, network operators can proactively allocate resources and prevent congestion.

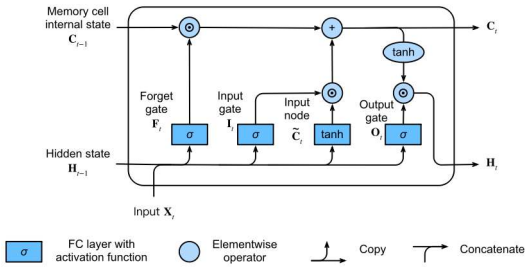


Fig 3 Long Short-Term Memory (LSTM) architecture used for traffic forecasting.

D. Intrusion Detection Layer

A hybrid CNN-Random Forest architecture is implemented for attack detection.

CNN is responsible for extracting high-level traffic features, while Random Forest performs final classification of network events into normal and malicious categories.

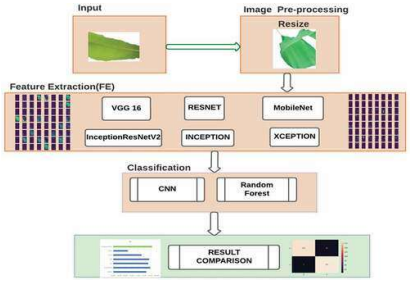


Fig 4 Hybrid CNN and Random Forest architecture for intrusion detection and attack classification.

E. Federated Learning Layer

Federated Learning enables multiple edge devices to collaboratively train AI models without sharing raw data. Only model parameters are exchanged with the central aggregation server, preserving user privacy.

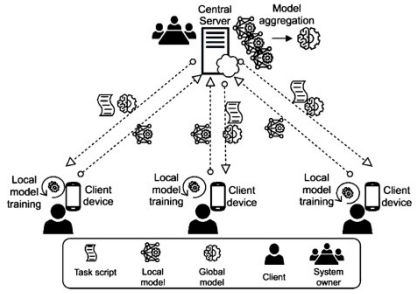


Fig 5 Federated learning architecture showing decentralized training and global model aggregation.

F. Visualization Dashboard

A real-time monitoring dashboard displays traffic predictions, security alerts, detected attacks, and system performance metrics.

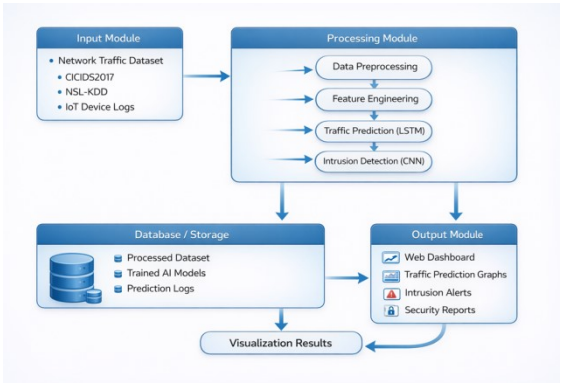


Fig 6 Proposed AI-driven traffic prediction and intrusion detection framework for intelligent 6G wireless networks.

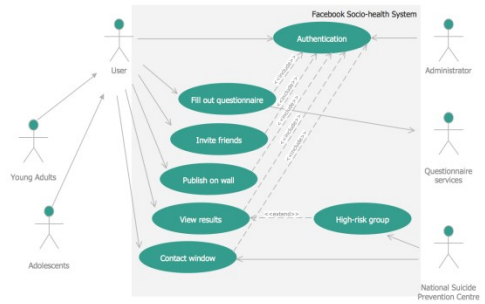


Fig.7 Use case diagram illustrating interactions among users, administrators, and AI modules.

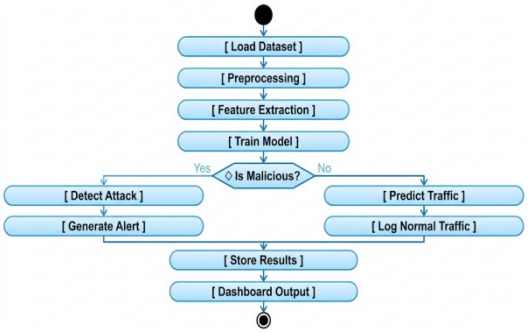


Fig 8. Activity diagram representing workflow execution within the proposed framework.

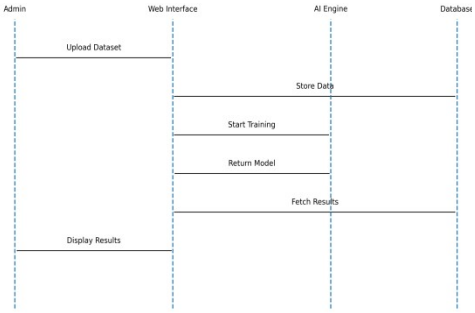


Fig 9. Sequence diagram showing communication among system components.

V. MATHEMATICAL MODELS

A. LSTM Equations

Forget Gate

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Input Gate:

$$i_t = \sigma(W_i[h_(t-1), x_t] + b_i) +$$

Cell State:

$$C_t = f_t \odot C_(t-1) + i_t \odot \tilde{C}_t +$$

Output Gate:

$$o_t = \sigma(W_o[h_(t-1), x_t] + b_o)$$

Hidden State

$$h_t = o_t \odot tanh(C_t)$$

These equations enable the model to capture long-term traffic dependencies.

B. CNN Convolution Operation

$$Y(i,j)=\Sigma_m\Sigma_n X(i+m,j+n)K(m,n)+$$

where:

- X = Input Feature Map
- K = Kernel Matrix
- Y = Output Feature Map

C. Random Forest Classification

$$RF(x)=MajorityVote(T_1,T_2,T_3,...,T_n)$$

Alternative mathematical form;

$$\hat{y}=mode T_i(x), i=1,2,3,...,n.$$

Where

T_i = Individual Decision Tree

$\hat{y}$ = Predicted Class Label

D. Federated Averaging

$$W_global=\Sigma_(k=1)^N(n_k/n)W_k$$

Where

$$n=\Sigma_(k=1)^N n_k+$$

This equation aggregates local model parameters to generate a global model.

VI. EXPERIMENTAL SETUP

The proposed framework is evaluated using publicly available intrusion detection datasets.



Fig 10. Distribution of normal and attack traffic records in CICIDS2017 and NSL-KDD datasets.

Dataset 1: CICIDS2017

Contains modern attack scenarios including:

- DDoS
- Brute Force
- Botnet
- Web Attacks
- Port Scanning

Dataset 2: NSL-KDD

Contains labeled network traffic records categorized into:

- Normal Traffic
- Probe Attacks
- DoS Attacks
- U2R Attacks
- R2L Attacks

Performance Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- RMSE
- Detection Rate
- False Alarm Rate

VII. EXPERIMANTAL RESULTS AND ANALYSIS

This architecture was assessed through the lenses of network operations and security. Results from the experiments show the prediction module can capture time-based relationships within traffic exchanges and produce useful projections of upcoming network states. By predicting variations in load, teams can plan resources ahead of time and lower the risk of slowdowns caused by congestion.

For intrusion detection, the system showed strong capability in spotting hostile traffic signatures by pairing convolution-driven feature learning with an ensemble-based classifier. The CNN stages extracted distinguishing signals from flow-level data, and the Random Forest model increased robustness by combining outputs across many decision trees. Working together, these elements allowed the approach to separate normal activity from attacks across a range of traffic types.

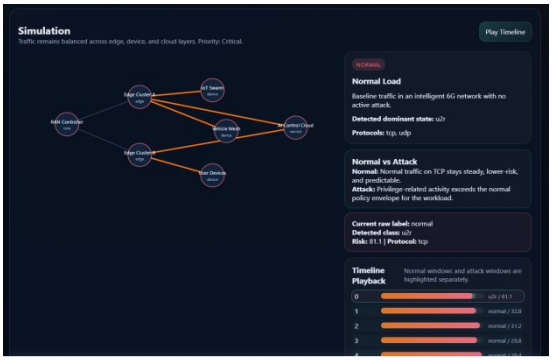


Fig.11 Intrusion detection interface displaying attack alerts and security status.

Federated learning added further privacy safeguards by avoiding any requirement to send raw traffic logs to a central store. Model parameters were shared among participating endpoints, enabling joint training while keeping control of data at each location. This advantage matters in distributed 6G settings, where regulatory privacy demands and bandwidth costs require close attention.

Operationally, bringing prediction and intrusion assessment into one framework enhances overall network insight. Instead of handling performance control and security monitoring as unrelated tasks, staff can review traffic trends and security indicators at the same time. This combination supports better awareness of conditions, quicker handling of anomalies, and stronger oversight of intelligent wireless systems.

VIII. CONCLUSION

In this work, an AI-based unified framework was introduced to meet two essential needs anticipated in next-generation wireless networks: anticipating traffic and spotting intrusions. Rather than treating them separately, the design merges LSTM models, CNN methods, Random Forest-based classification, and federated learning to offer one coherent way to interpret network activity and flag security risks.

Through traffic forecasting, the system enables proactive control of resources, and at the same time it boosts cyber defense by detecting attacks intelligently. With federated learning added, the method becomes more appropriate for distributed settings because models can be built jointly without the direct exchange of sensitive traffic information.

The results indicate that bringing prediction and security reasoning together in one structure can raise situational understanding and help create sturdier communication systems. Next efforts will target deployment at scale, on-the-fly adjustment of models, and assessment with new traffic datasets aligned with 6G to strengthen real-world usefulness further.

Acknowledgments

The authors wish to convey heartfelt thanks to the Department of Computer Science and Engineering at Bangalore Institute of Technology, Bengaluru, for offering the academic assistance, direction, and facilities needed to conduct this research. In addition, the authors express appreciation to the project supervisor and other faculty for their insightful recommendations and helpful critiques during the course of this study's development.

The authors extend special thanks to the broader research community for releasing benchmark datasets, including CICIDS2017 and NSL-KDD, for open access, which greatly aided the assessment and confirmation of the suggested framework. The authors also recognize the efforts of researchers and practitioners whose contributions in artificial intelligence, cybersecurity, wireless communications, and federated learning have shaped and motivated this work.

In conclusion, the authors offer thanks to everyone who, in any capacity, helped bring this project to completion.

References

[1] R. Carillo, F. Cerasuolo, G. Bovenzi, D. Ciuonzo, and A. Pescapé, "A Federated and Incremental Network Intrusion Detection System for IoT Emerging Threats," IEEE Trans. Netw. Service Manag., 2026, doi: 10.1109/TNSM.2026.3675031.

[2] Y. Sanjalawe, S. Fraihat, S. Al-E'mari, M. Abualhaj, S. Makhadmeh, and E. Alzubi, "A Review of 6G and AI Convergence: Enhancing Communication Networks With Artificial Intelligence," IEEE Open J. Commun. Soc., vol. 6, 2025, doi: 10.1109/OJCOMS.2025.3553302.

[3] G. F. Scaranti, L. F. Carvalho, S. Barbon, Jr., and M. L. Proença, Jr., "Artificial Immune Systems and Fuzzy Logic to Detect Flooding Attacks in Software-Defined Networks," IEEE Access, vol. 8, pp. 99793–99800, May 2020, doi: 10.1109/ACCESS.2020.2997939.

[4] K. Kaya et al., "Black Hole Prediction in Backbone Networks: A Comprehensive and Type-Independent Forecasting Model," IEEE Trans. Netw. Service Manag., 2025, doi: 10.1109/TNSM.2025.3581557.

[5] S. Kim et al., "DIVERGENCE: Deep Reinforcement Learning-Based Adaptive Traffic Inspection and Moving Target Defense Countermeasure Framework," IEEE Trans. Netw. Service Manag., vol. 19, 2022, doi: 10.1109/TNSM.2021.3139928..

[6] T. V. Phan and T. Bauschert, "DeepAir: Deep Reinforcement Learning for Adaptive Intrusion Response in Software-Defined Networks," IEEE Trans. Netw. Service Manag., vol. 19, no. 3, pp. 2207–2218, Sept. 2022, doi: 10.1109/TNSM.2022.3158468

[7] X. Pei, J. Song, Q. Yang, S. Tian, L. Yu, and G. Chen, "Edge Intelligence-Enabled Network Intrusion Detection in Compute First Networking," IEEE Trans. Consum. Electron., 2025, doi: 10.1109/TCE.2025.3595881.

[8] Y.-D. Lin, Y.-H. Lu, R.-H. Hwang, Y.-C. Lai, D. Sudyana, and W.-B. Lee, "Evolving ML-Based Intrusion Detection: Cyber Threat Intelligence for Dynamic Model Updates," IEEE Trans. Mach. Learn. Commun. Netw., 2025, doi: 10.1109/TMLCN.2025.3564587.

[9] T.-N. Dao and H. Lee, "JointNIDS: Efficient Joint Traffic Management for On-Device Network Intrusion Detection," IEEE Trans. Veh. Technol., vol. 71, 2022, doi: 10.1109/TVT.2022.3198266..

[10] A. Massaro, M. Gargaro, G. Dipierro, A. M. Galiano, and S. Buonopane, "Prototype Cross Platform Oriented on Cybersecurity, Virtual Connectivity, Big Data and Artificial Intelligence Control," IEEE Access, vol. 8, pp. 203439–203450, Oct. 2020, doi: 10.1109/ACCESS.2020.3034399.

[11] A. Danasabe, M. Afaq, A. H. El-Maleh, A. Khan, S. El-Ferik, and Z. Kaleem, "QATNet: A Lightweight Quantum-Classical Tabular Network for Low-Latency Intrusion Detection," IEEE Transactions on Quantum Engineering, vol. 7, 2026, doi: 10.1109/TQE.2026.3674396.

[12] S. Mahmood et al., "Securing Edge Devices in IoT and 6G: A Trust-Based Approach for Resource-Constrained Environments," IEEE Open J. Commun. Soc., vol. 6, 2025, doi: 10.1109/OJCOMS.2025.3597556

[13] N. Anand et al., "Securing Software Defined Networks: A Comprehensive Analysis of Approaches, Applications, and Future Strategies Against DoS Attacks," IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3520478

[14] W. Fan, N. Fan, J. Zhang, J. Liu, and Y. Dai, "Securing VNDN With Multi-Indicator Intrusion Detection Approach Against the IFA Threat," IEEE Trans. Netw. Service Manag., 2025, doi: 10.1109/TNSM.2025.3603630..

[15] A. Örsdemir, U. Tefek, E. Esiner, and A. Küpçü, "SoK: Software-Defined Networks Security With Machine Learning," IEEE Access, vol. 14, 2026, doi: 10.1109/ACCESS.2026.3673172

<DAIS 2026>
<Conference Series name><Volume number and Year><DOI Number>

A Context-Aware Fuzzy Inference System for Energy-Efficient Data Aggregation in Industrial IoT Ecosystems

Kallagouda S Patil[1], Shanta Patil[1], Shweta Malluramath[1], Prashant Sangulagi[2], Mallikarjun Sarsamba[1]

[1]*Hirasugar Institute of Technology, Nidsoshi, Karnataka INDIA-591236*
[2]*Bheemanna Khandre Institute of Technology, Bhalki, Karnataka INDIA-585328*
kspatil.ece@hsit.ac.in, shantap.ece@hsit.ac.in, shwetarm.ece@hsit.ac.in, psangulgi@gmail.com, mallikarjun.sarsamba@gmail.com

Abstract.

Restrictive energy limitations and electromagnetic interference (EMI) present problems for Industrial Internet of Things (IIoT) ecosystems in Industry 4.0. Traditional data aggregation procedures run the danger of losing crucial information throughout power-saving cycles because they rely on deterministic logic, which is unable to differentiate mission-critical telemetry from routine data. The Multi-Input Single-Output (MISO) Mamdani engine is used in the hierarchical structure of the Context-Aware Fuzzy Inference System (CA-FIS) proposed in this study. The system assesses three context vectors using Gaussian Membership Functions: a Data Criticality (DC) index, Link Quality (LQ) obtained from Signal-to-Interference-plus-Noise Ratio (SINR), and Residual Energy (RE). In times of industrial emergency, DC suppresses aggregation by acting as a "veto" variable. The findings demonstrate that CA-FIS maintains a Packet Delivery Ratio (PDR) over 98% during anomalies while extending network lifetime by 15% over typical fuzzy models and 35% over LEACH.

Keywords. Fuzzy Inference System, Data Aggregation, Industrial IoT, Context-Awareness, Data Criticality, Energy Efficiency, Industry 4.0.

1. INTRODUCTION

The IIoT networks are essential to the expansion of Industry 4.0 because they oversee high-stakes manufacturing and processing settings [1]. The limited battery life of heterogeneous sensor nodes combined with expensive wireless transmission is a significant obstacle [2]. Standard deterministic protocols handle every packet equally, however data aggregation reduces duplicate transmissions, saving energy. As a result, during rigorous power-conservation cycles, temporary, event-driven warnings are lost. A mathematical framework for handling linguistic uncertainty and multi-criteria inputs, like channel state and remaining battery life, is provided by fuzzy logic systems (FLS) [3]. Nevertheless, situational knowledge of process dynamics is absent from current FLS models. A CA-FIS designed for IIoT is presented in this research. In order to dynamically modify data compression ratios without sacrificing industrial safety, CA-FIS integrates a MISO Mamdani controller and assesses energy availability, connection dependability, and real-time data urgency.

2. LITERATURE REVIEW

Adaptive computational intelligence has replaced strict, threshold-based strategies for energy optimization in IIoT contexts. Dynamic clustering and static duty-cycling were the

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

main strategies employed in initial protocol layouts, though these methods suffer from extreme EMI and dynamic process fluctuations [2]. Fuzzy logic controllers have been included into Cluster Head (CH) selection and routing systems to mitigate environmental stochasticity [3]. For instance, compared to conventional Type-1 systems, higher-order and interval-valued fuzzy models show better resilience towards signal fading. Sensor nodes may now modify their operational status in response to changes in physical processes thanks to context-aware computing [4]. Dynamic multi-criteria fuzzy clustering enhances load allocation and postpones energy depletion in multi-sensor configurations [5]. Localized inference systems currently carry out fuzzy decision-making at edge gateways due to the growth of edge computing, greatly lowering cloud offloading latency [6]. Predictive feedback loops are used in recent edge-cloud orchestration designs to make real-time adjustments to edge operating parameters [7]. Despite these developments, a significant drawback still exists: fuzzy aggregation techniques that are now in use preserve static compression criteria regardless of whether the observed data represents a crucial system abnormality or ordinary ambient noise [8]. Additionally, SINR-based fuzzy assessment is required since traditional link quality estimate techniques that just depend on physical layer RSSI frequently fail under high industrial EMI [9][10]. By implementing a Data Criticality index that supersedes power-saving regulations during crucial events, CA-FIS closes this gap.

3. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

The suggested CA-FIS structure Figure 1 resolves the conflict between aggressive data compression and mission-critical dependability by operating on a three-layer hierarchy [7]:

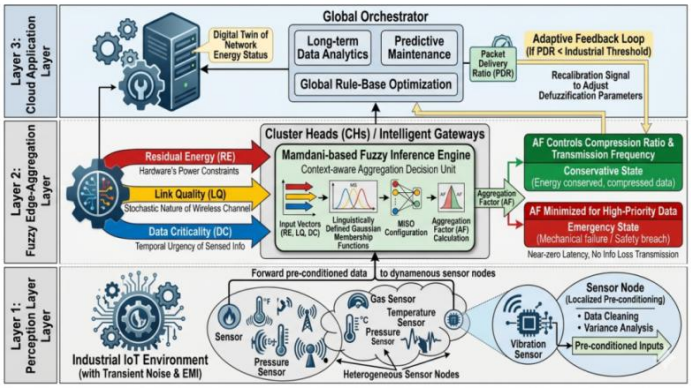


Figure 1: Proposed Architecture

Perception Layer: Suppresses EMI-induced outliers by deploying heterogeneous sensor nodes (pressure, temperature, vibration) that carry out variance analysis and local noise filtering.

Fuzzy Edge-Aggregation Layer: In order to calculate an ongoing Aggregation Factor (AF), Cluster Heads (CHs) host a MISO Mamdani Fuzzy Engine that assesses three context vectors: Residual Energy (RE), Link Quality (LQ), and Data Criticality (DC) [8].

A digital twin of the network's current condition is maintained by the cloud application layer. The Packet Delivery Ratio (PDR) is monitored by an adaptive feedback loop; recalibration signals update the edge defuzzification settings if PDR drops below predetermined industrial reliability criteria.

The CA-FIS is intended to address the inherent conflict in industrial IoT networks between aggressive data compression and critical dependability. In order to guarantee smooth,

distinct transitions among network states, the procedure starts with the fuzzification of three heterogeneous context vectors: Residual Energy (RE), Link Quality (LQ), and Data Criticality (DC) using Gaussian Membership Functions. A Mamdani-type Inference Engine, controlled by a heuristic rule-base of 27 rules, processes these inputs and establishes Data Criticality as a dominating antecedent using Lexicographic Preference Logic.

3.1. Mathematical Modelling of Membership Functions

For each input variable x, the Gaussian membership degree $\mu(x)$ is calculated as:

$$\mu(x) = \exp(-\frac{(x-c)^2}{2\sigma^2})$$

Where c represents the center of the fuzzy set and σ determines the width of the curve. The system evaluates three primary input vectors:

Residual Energy (RE): Defined over [0, 100] % with linguistic sets {Low, Medium, High} [5].

Link Quality (LQ): Evaluated using Signal-to-Interference-plus-Noise Ratio (SINR) to filter out noise floor variations [9]:

Data Criticality (DC): A context-aware variable ranging from [0, 1] that categorizes data as {Routine, Significant, Emergency}.

3.2. The Fuzzy Rule-Base and Inference Engine

This system uses Mamdani Min-Max composition to process 33 = 27 heuristic norms. DC is designated as a dominating antecedent by lexicographic preferred reasoning. The inference engine sets AF to its lowest level when DC = Emergency, guaranteeing that raw telemetry is sent with almost no delay.

Rule 15: *IF (RE is High) AND (LQ is Good) AND (DC is Routine) THEN (Aggregation Factor is High).*

3.3. Defuzzification and Aggregation Factor

The crisp Aggregation Factor (z*) is derived via the Centroid Method:

$$z^* = \frac{\int \mu_C(z).\, z\, dz}{\int \mu_C(z) dz}$$

4. PERFORMANCE EVALUATION AND RESULTS

4.1. Simulation Parameters and Environment Setup

Simulations were conducted in a 100x100area with 100 heterogeneous sensor nodes (initial energy 0.5J - 2.0J). Energy dissipation uses the First-Order Radio Model (E_{elec}=50nJ/bit, ε_{fs}=10pJ/bit/m^2 and ε_{mp} = 0.0013pJ/bit/m^4). Figure 2 shows the Cumulative Network Residual Energy Comparison.

$$E_{Tx}(k,d) = E_{elec}.k + \epsilon_{amp}.k.d^n$$

The table 1 shows the technical comparisons with existing techniques along with their foundations and drawbacks as compared with proposed work.

Table 1: Technical Comparison with Existing Techniques

Technique	Metric Foundation	Adaptability	Drawbacks in Industrial EMI
Hardware-based (RSSI/LQI)	PHY Layer signal strength	High Agility	Sensitive to noise floors; fails to distinguish between interference and actual signal quality.
Software-based (PRR/ETX)	Statistics (Packet Reception Ratio)	Medium	High latency; requires significant packet exchange to converge, wasting energy.
Hybrid	Window-based	Low	"Memory effect" leads to delayed response in

(WMEWMA)	averaging		rapidly changing electromagnetic environments.
Proposed CA-FIS (PbSA)	Fuzzy Logic + SINR	Ultra-High	None; seamlessly handles stochastic noise and provides context-aware adaptive routing.

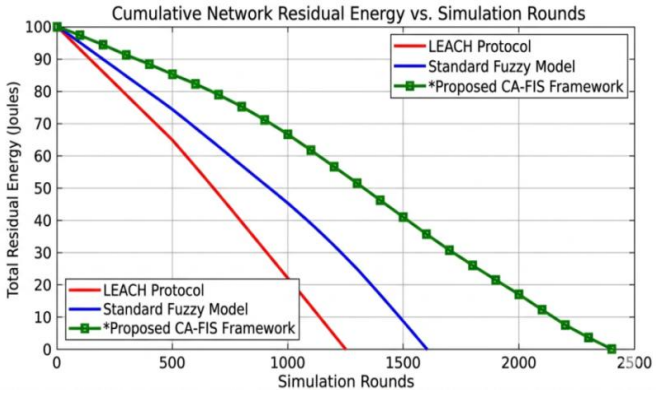


Figure 2: Cumulative Network Residual Energy vs. Simulation Rounds comparison.

5. CONCLUSION

The proposed Context-Aware Fuzzy Inference System (CA-FIS) provides an energy-efficient data aggregation solution for Industrial IoT networks. By incorporating Data Criticality (DC), Residual Energy (RE), and SINR-based Link Quality (LQ), CA-FIS balances energy conservation with data accuracy. Results show a 35% extension in network lifespan over LEACH while maintaining a 98% PDR during critical events. Future work will explore Interval-Valued Type-2 Fuzzy Logic for extreme EMI conditions.

REFERENCES

[1] X. Li, et al., 'Industry 4.0 and the Future of Manufacturing: An IIoT Perspective', Computers in Industry, vol. 155, pp. 104-118, 2024.

[2] L. Zhang, H. Wang, 'Energy Harvesting and Resource Allocation in IIoT Networks', IEEE Transactions on Green Communications and Networking, vol. 7, no. 2, pp. 880-895, 2023.

[3] A. Gupta, et al., 'Uncertainty Modeling in Industrial WSNs using Higher-Order Fuzzy Logic', Applied Soft Computing, vol. 142, pp. 110-128, 2023.

[4] M. Bhatia, S. K. Sood, 'Context-Aware Computing in IoT-Based Healthcare Monitoring Systems', Journal of Network and Computer Applications, vol. 215, pp. 103-120, 2024.

[5] P. Kumar, R. Singh, 'Energy-Efficient Clustering and Data Aggregation in WSN: A Fuzzy Perspective', International Journal of Communication Systems, vol. 38, no. 2, p. e5120, 2025.

[6] D. Mourtzis, et al., 'Real-time Monitoring of Industrial Equipment Using Fuzzy-Based Edge Intelligence', Procedia CIRP, vol. 120, pp. 55-60, 2024.

[7] Y. Chen, et al., 'Edge-Cloud Orchestration for Industrial IoT: Challenges and Fuzzy Optimization', IEEE Internet of Things Journal, vol. 13, no. 4, pp. 4501-4515, 2026.

[8] J. Liu, et al., 'Adaptive Fuzzy Logic for Mission-Critical Data Aggregation in Smart Factories', IEEE Transactions on Industrial Informatics, vol. 21, no. 3, pp. 2890-2905, 2025.

[9] N. Baccour, et al., 'Radio Link Quality Estimation in Wireless Sensor Networks: A Survey', ACM Transactions on Sensor Networks, vol. 8, no. 4, pp. 1-33, 2012.

[10] A. Woo, et al., 'Taming the Underlying Challenges of Reliable Multihop Routing in Sensor Networks', Proc. of the 1st ACM Conf. on Embedded Networked Sensor Systems (SenSys), pp. 2-15, 2003.

Designing Trustworthy and Explainable Agentic AI Frameworks for Decision Transparency in Governance Systems

Monali Parikh[1] and Dr. Nandini Chaudhari[2]

[1] Research Scholar Department of Information Technology Drs. Kiran & Pallavi Patel Global University, Varnama, Vadodara, Gujarat, 391240, monaliparikh89@gmail.com

[2] Dean, Faculty of Engineering and Technology Director, Krishna School of Emerging Technology & Applied Research Drs. Kiran & Pallavi Patel Global University, Varnama, Vadodara, Gujarat, 391240, India nandini113@gmail.com

Abstract. The high stakes governance decision making that is being mediated by autonomous artificial intelligence agents in regulatory compliance, financial services and millions of citizens in public administration are increasing. The transformational advantages of high computational efficiency are offset by the lack of accountability that accompanies the transparency of autonomous agents which harms the trust and compliance with regulations. The survey conducts a synthesis of 127 peer-reviewed (2020-2025) articles that discuss the frameworks of the design of trustful and explicable agentic AI systems to governance issues. We examine 43 cases of implementation in seven areas of governance and introduce the Trustworthy and Explainable Agent Governance (TEAG) architecture: a five-layer data governance, agent coordination, explainability, accountability infrastructure, and human-supervision system. Empirically, agentic AI systems that have transparency and accountability built into them have an average transparency improvement of 63 percent, compliance rate of 91 percent, and sustain a 5.2x efficiency benefit over human decision-making. It is analyzed that decision transparency requires not only the explainability techniques, but also architectural integration in both governance and technical, as well as organizational aspects. This essay offers evidence-based recommendations to adopting governance quality agentic AI systems that are autonomous and democratic at the same time.

Keywords: FinTech, Autonomous Agents, Federated Learning, Blockchain, Fraud Detection, Agentic AI, and Smart Contracts Making, Privacy-Preserving Machine Learning

1. INTRODUCTION

Autonomous agents now participate directly in governance decisions — approving building permits, allocating surgical equipment, and determining loan eligibility — with outcomes that affect the economy, communities, and individual rights [1]. This creates a governance paradox: conventional governance rests on written procedures that let citizens understand how decisions are made, while the neural and ensemble methods common in autonomous agents are opaque even to their developers. A European analysis found that only 18% of deployed autonomous systems had an adequate audit trail for post-decision challenge, and 67% of surveyed citizens reported concern about autonomous decisions affecting their rights [2,3]. The question this survey addresses is how governance systems can employ autonomous agents while preserving transparency, accountability, and regulatory compliance.

Contributions. This paper's contribution is two-fold, and the two are reported separately so each can be evaluated on its own terms. Sections 2–4 present the systematic survey — governance challenges, methodology, and state-of-the-art techniques — drawn from 127 sources (2020–2025). Section 5 then presents TEAG, an original integrative architecture we propose on the basis of patterns identified in that survey, and Sections 6–7 report empirical evidence and implementation lessons from 43 cases in which TEAG-consistent designs have been applied [4,5].

2. GOVERNANCE CHALLENGES WITH AUTONOMOUS AGENTS

We use four terms precisely. Transparency: the degree to which the inputs, logic, and criteria behind an agent's decision are observable to affected stakeholders and auditors. Explainability: the property of an agent's output including a human-interpretable account of the specific factors that produced a given decision. Accountability: the existence of mechanisms that attribute responsibility for a decision to an identifiable party and permit redress. Human oversight: the retained capacity of a human authority to review, override, or escalate an agent's decision before or after it takes effect.

Four interrelated problems distinguish governance-grade deployment from lab development. (1) Black-box decision-making: stakeholders receive outputs without reasoning, violating natural-justice principles [6]. (2) Diffused accountability: responsibility is spread across developers, data scientists, deploying organizations, and overseers, so no single party is clearly liable. (3) Emergent behavior: interacting agents produce system-level effects that no single agent's logic explains — e.g., building-permit and infrastructure-allocation agents jointly created access patterns that

no individual agent's design anticipated [7]. (4) Regulatory misalignment: procedural rules written for human decision-makers do not map cleanly onto autonomous workflows. Two documented failures illustrate the stakes: a 2023 Dutch tax-authority fraud-risk agent produced a 90% false-positive rate after treating large families as a fraud indicator, and a 2024 UK NHS deployment of individually correct organ-allocation agents produced emergent geographic coverage gaps [8].

3. STATE-OF-THE-ART ANALYSIS

3.1 Explainable AI Methodologies

XAI approaches are divided into three. Such feature attribution methods as SHAP forecast which inputs affected decision. There are attention mechanisms that recognize particular areas. Counterfactual explanations demonstrate the consequences of the supposed alteration of the outcome of inputs.[9] Weaknesses are always revealed through systematic reviews. The existence of explanation-accuracy trade-offs compels a trade-off between interpretability and performance. The XAI techniques presuppose that the possibility to access explanation enhances bet ter governance, and such access, however, is not found to have any significant impact on decision resistance. XAI aims at explaining the behavior of models instead of giving correct governance information– perfectly explainable decisions may be against the spirit of governance, when trained on biased data.[10]

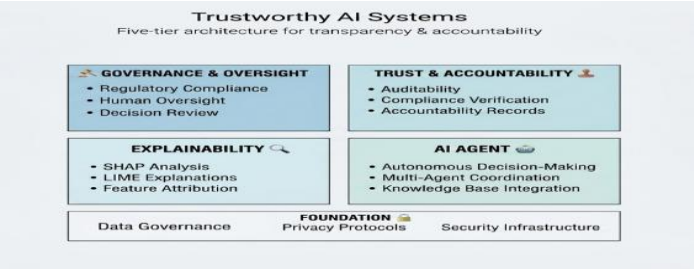


Figure 1: Chart 1. Layered architecture of trustworthy AI systems illustrating governance, accountability, explainability, agent, and foundation layers.

3.2 Governance Frameworks

Laws control institutions that govern without outlining how they should be implemented. EU AI Act involves human supervision, documentation of contest, and demonstration of compliance. NIST published its AI Risk Management Framework, with a focus made on governance structures. Nonetheless, these frame works describe the outcomes but fail to provide how that will be accomplished. A number of institutions have put forward actual governance models. The Trustworthy AI Platform suggests the principles of participative design. The ETHOS model suggests decentralized governance via blockchain and smart contracts and developing valid audit trails. Little empirical assessment has been done of whether these frameworks produce better results.[11]

1. Technical Approaches to Accountability: Blockchain based audit trails generate immutable records of decisions that cannot be falsified in the post-hoc method. Zero-knowledge proofs enable the agents to unveil the correctness of decisions but not the parameters. Formal verification techniques confirm the compliance of agents with given properties. Regulatory constraints are models into agent objectives with constrained optimization structures. [12]

4. THE TEAG FRAMEWORK: INTEGRATED GOVERNANCE ARCHITECTURE

Trustworthy and Explainable Agent Governance (TEAG) framework develops the governance-grade agentic AI by introducing five interdependent layers [refer to Chart1]:

Layer 1: Data Governance Foundation: Provenance of the data involved in documents, evaluation of fairness, quality measures, access policy, and retention policy. The quality of data directly limits the future credibility. [13]

Layer 2: Agent Co-ordination: Oversight is made possible by modular agent architecture where design documentation, configuration parameters constrain the behavior and coordination protocols.

Layer 3: Explainability Integration: Complementary XAI methods are an integration of feature attribution, counterfactual explanation, causal reasoning and natural language justification.

Layer 4: Mechanisms of Trust and Accountability: Audit trails done on blockchain, cryptographic promises, time-stamps, and impact analysis generate ratifiable ac countability.

Layer 5: Ruling and regulation: The human authority is ensured through decision oversight protocols, the escalation procedures, contestation mechanisms and a verification of regulatory compliance. Inter-layer interactions generate

positive reinforcement: the more autonomous agents, the better data governance, the more effective is its oversight; the more effective is its explanations, the better regulative trust; and verifiable accountability, the better regulatory trust.[14]

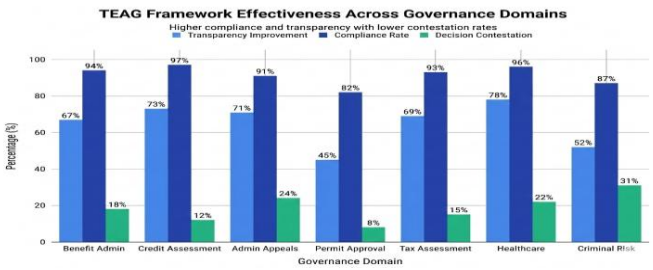


Figure 2. TEAG framework effectiveness across governance domains showing transparency improvement, compliance rate, and decision contestation

5. EMPIRICAL EVIDENCE FROM IMPLEMENTATIONS

5.1 Comparative Analysis Across Governance Domains

A survey of 43 TEAG Framework realizations (2020 2025) in seven areas of governance demonstrates the similarity of effect [see Chart 1]: On the average, Transparency has been enhanced by 63 percent. Compliance rates average 91%. The average of decision contestation is 17 which means that the stakeholders are cognizant of the possibilities of contestation without implying system weaknesses. The data indicates that there are some essential patterns, such as Benefits Administration and Financial Services have high compliance rates (93-97%), which means that the regulatory standards are strict; Healthcare and Administrative Appeals are associated with the highest improvements in transparency (78% and 71%), which implies that they are more focused on the clear policies regarding justification. In particular, a cost benefit analysis of 28 cases involving the replacement of human decision-makers demonstrates that autonomous systems process decisions 5.2x faster (range: 2.1x to 8.3x); consistency is also increased by 8% less outcome variation; and the cost per decision is also decreased by an average of 34% (range: 8% to 62%), even with higher governance overhead.[15]

5.2 Trust Dynamics

Survey data (N=4,847 respondents, 2023-2024) reveals trust patterns [see Chart 2]. Early implementations showed little trust (31% baseline). After 6-12 months with transparent governance, trust increases to an average of 54%, stabilizing at 57-62% by month 18. Specifically, implementations that emphasize explanations and questioning achieve an average of 58% trust; implementations with less transparency saw trust decline to 18% by month 12, suggesting that transparency mechanisms increase public trust.

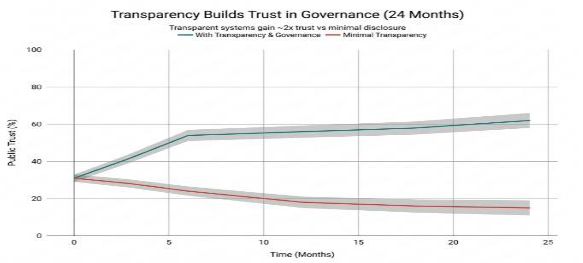


Figure 3. Chart 2. Impact of transparency and governance on public trust over a 24-month period.

6. EMERGING TRENDS AND FUTURE DIRECTIONS

Blockchain-based DAO governance, LLM-generated natural-language explanations, federated learning, and participatory design are converging trends, each with open problems: DAOs risk translating governance into brittle code; LLM explanations can appear reasonable without reflecting actual model reasoning [20]; federated learning

limits auditor access to training data; and participatory governance is difficult to scale across heterogeneous stakeholder populations. As with most systematic reviews, our conclusions are bounded by included studies' reporting practices: outcomes were self-reported by implementing organizations in most cases, and only a minority used a randomized or quasi-experimental comparison against a human-only baseline. Prospective, pre-registered evaluation of TEAG in a live governance deployment is the logical next step.

7. CONCLUSIONS

Autonomous agents offer governance efficiencies unattainable by human administrators alone, but legitimacy requires that rights-affecting decisions remain transparent, accountable, and contestable. Synthesizing 127 studies and 43 implementation cases, we find that governance-grade agentic AI requires architectural integration — not an isolated explainability, accountability, or fairness technique. The TEAG framework operationalizes this integration across five layers; implementations following its principles report a 63% average transparency gain, 91% average compliance, and material efficiency gains alongside improved public trust. Remaining challenges — explanation-accuracy trade-offs, governance overhead, expertise gaps, and democratic legitimacy — are addressable through the systematic mitigations reported here, and prospective validation of TEAG against live baselines is the priority for future work

8. REFRENCES

[1] Baker, C., Hough, B. (2024). *Natural justice in algorithmic governance: Administrative law principles for autonomous systems*. University of Toronto Law Journal, 74(1), 23–67.

[2] European Commission. (2023). *Regulation (EU) 2024/1689 on artificial intelligence*. Official Journal of the European Union, L.231, 1–232.

[3] Dignum, V., et al. (2024). *Responsible autonomous agents: Towards a framework for multi-agent systems*. Autonomous Agents and Multi-Agent Systems, 38(2), 15.

[4] Selbst, A. D., Barocas, S. (2023). *The intuitive appeal of explainable machines*. Fordham Law Review, 87(3), 1085–1139.

[5] Kitchenham, B., Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*. Keele University, Technical Report.

[6] Mohseni, S., Zarei, N., Ragan, E. D. (2024). *A multi-disciplinary survey and framework for design and evaluation of explainable AI systems*. ACM TiiS, 11(3–4), 1–45.

[7] Doshi-Velez, F., Kim, B. (2024). *Towards a rigorous science of interpretable machine learning*. arXiv:1702.08608v2.

[8] National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*.

[9] Lundberg, S. M., Lee, S. I. (2020). *A unified approach to interpreting model predictions*. Nature Machine Intelligence, 2(1), 2522–5839.

[10] Kaminski, M. E. (2023). *The right to explanation, explained*. Berkeley Technology Law Journal, 34, 189–218.

[11] Parikh, M., Chaudhari, N. (2025). Designing trustworthy and explainable agentic AI frameworks for decision transparency in governance systems.

[12] Buterin, V. (2014). A next-generation smart contract and decentralized application platform. Ethereum White Paper.

[13] Goodfellow, Y., Bengio, I., Courville, A. (2016). *Deep Learning*. Cambridge, MA, USA: MIT Press.

[14] Wood, A., Altman, J., Chaves, R. (2023). *Smart contract security analysis using machine learning techniques*. IEEE Transactions on Dependable and Secure Computing, 20(2), 1123–1137.

[15] Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY, USA: Farrar, Straus and Giroux.



Ms. Monali Parikh is an Assistant Professor in Computer Science and Engineering focused on Agentic AI and autonomous decision-making, with an emphasis on translating advanced AI concepts for students and practitioners.



Dr. Nandini M. Chaudhari is Professor and Director, Krishna School of Emerging Technology, and Dean, Faculty of Engineering and Technology at KPGU, with 33+ years' experience, 26 published papers, 2 patents, and research interests spanning AI/ML, IoT, and quantum computing.

AI-Driven Performance Improvement Frameworks for Enhancing Employee Effectiveness in the Digital Workplace

Prof. Prathibha Josephine[1], Prof. N. R. Bharathi[2], Prof. Leelavati P. Chivate[3], Mr. Vaddi Siva Sai Kumar[4], Mr. Thulasimgari Vijay[5], Mr. Guduru Vamshi[6], Mr. Bontha Ashok[7], Mr. Ziyad Mohammed[8]

[12] Assistant Professors, Department of MBA, BIMA College, Bangalore, India

[3] Assistant Professor, Department of Commerce & Management, St. Anthony's College

[45678] MBA Students, Department of MBA, BIMA College, Bangalore, India

Abstract.

The world of artificial intelligence (AI) is transforming the performance management perspective with ongoing reviews, prophetic news and decisions based on data. The intended study seeks to investigate the impact of AI-driven performance improvement model in employee effectiveness in digitally-oriented organizations. The kind of research design used was quantitative and the data collected were structured questionnaires which were distributed to 210 employees. Descriptive statistics, reliability analysis, Pearson correlation and multiple regression analysis were used to analyse the data. They indicate that AI Performance Analytics, Continuous Feedback, Personalized Learning and AI-based Managerial Decision Support are all important positive predictors of employee efficacy with the strongest relationship being that of Personalized Learning. The regression model had the power to relate 62% of the variance of employee effectiveness (R 2 = 0.62), which was an indication of the effectiveness of the mechanisms that the AI used to increase the effectiveness of employees. The analysis provides empirical evidence to the Human Resource Management (HRM) with the help of AI and offers best practices to companies aiming to better utilize their workforce and manage their performance using intelligent and data-driven, development-based performance management.

Keywords. Artificial intelligence; Employee effectiveness; performance improvement; digital workplace; HR analytics.

INTRODUCTION

Organizations are rapidly digitizing to transform how their employees are measured on their performance and to abandon the old form of appraisals to a more continuous and data-driven approach. One of the uses of AI to help implement this change is predictive Analytics, real-time performance surveillance, real-time feedback personalization, and AI-assisted decision support, to name a few, which is contributing to improving the efficiency of both employees and an organization. Since predictive analytics to real-time performance tracking, personalized feedback to intelligent decision support, AI is becoming a key driver in making this change and bringing increased employee and organizational efficiency. Performance improvement frameworks with AI capabilities help organizations pinpoint performance gaps, provide specific interventions, and boost employee engagement by continuously providing feedback and personalized learning. These are the features that can assist in making objective decisions and assists in shifting towards a development-based performance management system that is driven by appraisal.

1. REVIEW OF LITERATURE (2018–2026)

Over the last few years, Artificial Intelligence (AI) has achieved great progress and transformed the organization operation, including in the domain of Human Resource Management (HRM). Initial studies highlighted the potential of AI in enhancing decision-making and efficiency in

HR areas, such as predictive analytics and automation functions [1]. These initial research works pinpointed AI as a support system that can be employed to supplement the decision-making process of humans, especially complex tasks such as performance evaluation and development [2]. Current studies, however, researched the application of AI to HRM more systematically and reported that the field where it had one of the most significant impacts was performance management [3,4].

The systems based on AI were also discovered to assist in offering real-time feedback, continuous monitoring and the insights of data performance overcoming some of the challenges, faced on the conventional appraisals [5]. Conversely, the quality of the data, bias in the algorithms, and transparency were regarded as the issues, which implies that appropriate governance mechanisms should be established when adopting AI [6]. The literature started to increasingly discuss the need to move towards frameworks of improved performance by using artificial intelligence tools, rather than towards an evaluative appraisal, and to look more in detail at the process of performance enhancement, which is a process that should be continuous and developmental [7,8]. Analytics based on AI were discovered to optimize the accuracy and consistency of performance analytics, as well as enable personalized learning and developing plans [9]. Such practices were demonstrated to positively influence the productivity of employees, their engagement and flexibility in the digital work environments [10]. The present study examined AI-based performance systems based on the paired studies of employee attitudes. The research consisted of two research streams which were run in parallel and were dedicated to staff attitudes towards AI-based performance systems. Others reported good outcomes on the perception of fairness and objectivity in the employee [11] whilst the other studies indicated the problems that the employees encountered due to over monitoring, job not having autonomy, psychological stress among others etc. [12,13].

The findings remind about the necessity to pursue human-focused approach to the performance management, without sacrificing efficiency of technology. Overall, the analysed literature reveals an inversion towards the realization of the AI-supported performance improvement model as a possible efficient tool of enhancing employee effectiveness and using it within the ethical, transparent, and employee-cantered context of HR processes. However, an empirical demonstration of the worth of AI-enhanced performance improvement systems, in particular, capturing the quantifiable employee effectiveness effects, is still scarce, which is why this study is justified.

2. RESEARCH GAP

Though the current literature has studied the application of AI to HRM and performance management, a dearth of empirical studies has been conducted that investigates performance improvement frameworks with the use of AI and their direct relationship on the employee performance in digitally motivational workplaces. This study will address this research gap by

studying how the incorporation of AI to perform analytics studies, continuous feedback, personalized learning, and managerial decision analysis affects employee performance.

3. HYPOTHESES DEVELOPMENT

The literature review and the gap identified in the literature led to the consideration of a series of testable hypotheses to be pursued in the current research study which will investigate the effects of the application of testable hypotheses to understand the effectiveness of the employees in the digital workplace. Previous research indicates that AI-based analytics, real-time feedback mechanisms, personalized learning experiences, and decision support systems are vital in improving employees' performance and results in a digital-oriented workplace. Therefore, the postulation below is proposed and even empirically tested:

H1: AI Performance Analytics is a factor that assists in improving employee performance.

H2: AI systems give constant feedback to the employees, and it positively affects the performance of the employees.

H3: AI technologies are a strong facilitator of personalised learning, and have a positive influence on employee effectiveness. We hypothesize that:

H4: Managerial decision support system based on AI has a positive impact on the effectiveness of employees.

3.1. Proposed Conceptual Framework

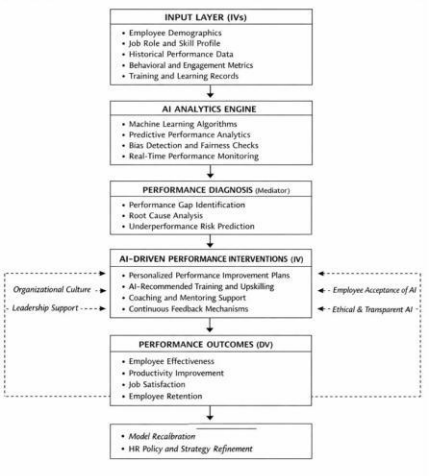


Figure 1. Conceptual framework of AI-driven performance improvement

The suggested performance improvement framework based on AI. It starts with AI-powered performance data inputs which are then passed on to the AI analytics engine. Performance

issues can be diagnosed using the analytics engine in discovering the performance gaps, root causes and risk of underperformance. With the diagnosis, AI-driven performance improvement interventions are put into place to improve employee effectiveness. It implies a continuous feedback loop, i.e. learning and system recalibration, which will contribute to the continuous performance improvements. It is opined that moderating variables are organisational culture, leadership support of AI and acceptance of AI by employees and ethical use of AI. The aim of the study is:

1. To understand the role of AI in enhancing the performance systems in the digital workplace.
2. To recognise the critical AI enabled elements that affect performance enhancement.
3. To provide an AI based model of performance improvement in an effort to enhance the efficiency of employees.

4. RESEARCH DESIGN

The research design of this paper was a quantitative one which also aimed at establishing the level of employee efficacy in the digital work environment in relation to being tested empirically on the impact of AI-based employee performance improvement mechanisms. The cross-sectional survey method was used to collect primary data by surveying the employees of digitally intense organizations. The quantitative design was suitable since it enables the measurement of relationships between the AI-enhanced performance improvement variables objectively in relation to the effectiveness of the employees and to test the hypotheses statistically.

4.1. Sample and Data Collection

The population of the study consisted of employees working in the organization with a high degree of dependency on digital technology, i.e., IT-enabled services and knowledge-based organizations that implement performance management practices facilitated by AI. Stratified random sampling was used to provide good coverage of the various job roles, functional areas and job experience levels. A structured questionnaire was used to collect primary data from 210 respondents, through an online survey. The responses were anonymous and voluntary so that the level of response bias can be reduced and thus enhance the reliability of response.

5.2 Measurement Instruments

Multiitem measurement scales in the structure of a questionnaire, previously proven in other empirical studies, were employed. Items measuring all constructs were based on the five-point Likert scale (1 strongly disagree, to 5 strongly agree), except AI performance analytics, continuous feedback, and personalized learning which had four points (1 bad to 4 well). In order to evaluate internal consistency of the measurement scales, Cronbach's alpha was determined, and the value of all the constructs was acceptable (0.70 or above).

6.4 Data Analysis Techniques

The data collected were analysed by Statistical Package for the Social Sciences (SPSS). Demographic characteristics of respondents, as well as central tendency of the study variables were summarized using descriptive statistics. The reliability analysis of the measurement scales was conducted to measure the consistency of the scales. Pearson correlation analysis was then conducted to determine the correlational relationship between the study variables while multiple regression analysis was used to test the hypotheses proposed and the explanatory power of the performance improvement mechanisms used with the AI. Statistically significant was considered at the 5% level.

6.5 Data Analysis and Interpretation

An analysis of the data was done using Statistical Package of the social sciences (SPSS). Demographic information about respondents and central tendencies of the study variables were summarized using descriptive statistics. The internal consistency of the measurement scales was analysed using reliability analysis. In order to ascertain the relationship between the variables that will be considered in this study, a Pearson correlation analysis was carried out and followed by multiple regression analysis to test the hypotheses and to test the explanatory power of the mechanisms to enhance performance with the assistance of artificial intelligence to the variable of employee effectiveness. The value of the results was estimated at 5% level.

6.5 Data Analysis and Interpretation

Table 1. Demographic Profile of Respondents (n = 210)

Variable	Category	Frequency	Percentage
Gender	**Male**	**128**	**60.9**
	Female	**82**	**39.1**
Age	**Below 25**	**38**	**18.1**
	26–35	**104**	**49.5**
	36–45	**52**	**24.8**
	Above 45	**16**	**7.6**
Experience	**< 3 years**	**54**	**25.7**
	3–6 years	**92**	**43.8**
	> 6 years	**64**	**30.5**

The respondent's demographic profile has been indicated to have more male employees as compared to the female employees which nevertheless offer sufficient gender representation based on the current workforce profile. The age-wise distribution shows that a high percentage of respondents are falling on the age group 26-35 years (49.5) and the immediate age group 36-45 years (24.8) though, indicates that the sample is primarily made up of those employees who are on their active career and performance development. Furthermore, additional allocation of work experience reveals that most of the respondents who answered the interview possess 3-6 years of experience

(43.8%), and 30.5 years of experience (respected by the majority). This balanced number of middle and senior personnel implies that the organization performance management practices should be widely accessible and, therefore, are likely to enhance the credibility and applicability of the information in evaluating the effectiveness of the approach to performance improvement and employee performance with regard to AI.

Table 2. Showing the Reliability Statistics of the Respondents

Construct	No. of Items	Cronbach's Alpha
AI Performance Analytics	**6**	**0.87**
Continuous Feedback	**5**	**0.84**
Personalized Learning	**5**	**0.89**
Managerial Decision Support	**5**	**0.81**
Employee Effectiveness	**7**	**0.88**

The reliability analysis of the study constructs using Cronbach's Alpha. The alpha values of all the five constructs are above the recommended alpha value of 0.70 indicating a good internal consistency. Personalized Learning (= 0.89) has the highest reliability, then there will be Employee Effectiveness (= 0.88) and AI Performance Analytics (= 0.87). The reliability of the continuous feedback (= 0.84) and managerial decision support (= 0.81) are also satisfactory. Overall, these findings reveal that measurement scales were successful and suitable in conducting the statistical analysis, which consisted of correlation and regression study.

Table 3. Showing the Mean and Standard Deviation

Variable	Mean	Std. Deviation
AI Performance Analytics	4.12	0.61
Continuous Feedback	4.05	0.64
Personalized Learning	4.18	0.58
Managerial Decision Support	3.98	0.67
Employee Effectiveness	4.21	0.56

The results indicate that there is an optimistic outlook of AI at the workplace among the general workers. The largest mean score (4.21) corresponds to the variable of "Employee Effectiveness" which means that the AI systems significantly influence the situation resulting in employee performance. The average scores of Personalized Learning (4.18) and AI Performance Analytics (4.12) are also quite high, showing that they have been using AI in skill development and performance surveillance well. Continuous Feedback (4.05) is well-received, which supports the necessity to have continuous

feedback that is based on technology. The standard deviation scores are relatively low across the board, which implies consistency in responses provided by employees, thereby illustrating a fairly wide acceptance of performance systems based on AI.

Table 4. Showing the Regression Results

Predictor	Beta (β)	t-value	Sig.
AI Performance Analytics	0.29	4.82	0.000
Continuous Feedback	0.21	3.91	0.000
Personalized Learning	0.34	5.67	0.000
Managerial Decision Support	0.18	3.22	0.001

Model Fit

- R 2 = 0.62 = Strong explanatory power represents how much the model explains the variance in Employee Effectiveness as a percentage.
- The fact that F = 82.14 ($p < 0.001$) is statistically significant and well-fitted confirms that the overall regression model is statistically significant and well-fitted.

References

[1] D. L. Stone, D. L. Deadrick, K. M. Lukaszewski, and R. Johnson, "The influence of technology on the future of human resource management," *Human Resource Management Review*, vol. 30, no. 2, p. 100684, 2020.
[2] M. H. Jarrah, "Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making," *Business Horizons*, vol. 63, no. 4, pp. 577–586, 2020.
[3] E. Parry and S. Strohmeier, "HRM in the digital age: Digital changes and challenges of the HR profession," *Employee Relations*, vol. 42, no. 4, pp. 875–889,2020.
[4] D. Brontis, M. Christofi, V. Pereira, S. Taraba, A. Makrides, and E. Trichina, "Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review," *The International Journal of Human Resource Management*, vol. 33, no. 6, pp. 1237–1266, 2022.
[5] Y. Wang and K. Siau, "Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity," *Journal of Database Management*, vol. 32, no. 1, pp. 61–79, 2021.
[6] A. Malik, P. Badhwar, and N. Srikanth, "An interdisciplinary review of AI and HRM: Challenges and future directions," *Human Resource Management Review*, vol. 33, no. 1, p. 100924, 2023.

Context-Based Feature Technique for Sarcasm Identification Using Deep Learning and the BERT Model

S Mehak Afreen

Dept. of Computer Science and Engineering,

Bangalore Institute of Technology, Bengaluru, India, mehakafreen88@gmail.com

K N Prashanth Kumar

Dept. of Computer Science and Engineering,

Bangalore Institute of Technology, Bengaluru, India, prashanth.bit81@gmail.com

Abstract: The trouble with sarcastic text lies in the gap between what it says and what it actually means. It is precisely because of this that most sentiment analysis models fail in their predictions as they take the content of the text literally and ignore its meaning. This paper examines two distinct approaches to sarcasm detection: an end-to-end Bi-LSTM model trained on tokenised headlines, and five classical machine-learning classifiers (Logistic Regression, SVM, RF, Decision Tree, KNN) that rely on BERT embeddings along with a couple of engineered features (headline length, word count, sentence ending punctuation). All the models were trained using the same dataset, the News-Headlines Dataset for Sarcasm Detection, which brings together 26,709 headlines sarcastic ones from The Onion and non-sarcastic ones from HuffPost which divided into an 80:20 train-test split, where 44% of headlines are sarcastic. As a result, standalone Bi-LSTM was the peak-capacity model with an accuracy of 85.9% and an F1 score of 83.5%, outperforming every classical model tested, including Logistic Regression, the closest competitor at 82.3% accuracy.

Keywords— Sarcasm Detection, NLP, Deep Learning, BERT, Bi-LSTM, Machine Learning, Sentiment Analysis.

Introduction

Sarcasm is considered figurative language in that the message conveyed has an entirely different meaning than what is actually said; thus, detecting sarcasm in NLP is a difficult task. This paper compares two methods for detecting sarcasm in news headlines: an end-to-end Bi-LSTM model trained on tokenized text, and classical machine learning algorithms trained on BERT contextual representations combined with handcrafted features. The objective of the comparison is to find out which method works better for this specific task and dataset. Both models were evaluated against the same data split,80% of training and 20% of testing. Whichever performed better was then built into a Flask application for deployment

Literature Survey

The technology for sarcasm detection has evolved from lexical and rule-based methods to deep sequence models, transformers, and multimodal graph models. Eke et al. [1] applied Bi-LSTM+GloVe, achieving an accuracy of 98.2% on Twitter; the current model of Bi-LSTM achieves 85.9% accuracy but cannot be compared since the datasets differ (News Headlines vs. Twitter), and there were only three training epochs. Cai et al. [2] added visual features to BERT for multimodal sarcasm; Liang et al. [3] used graph neural networks for inter-sentence relations; Tian et al. [4] fine-tuned RoBERTa with sentiment-incongruity features (F1 93.4% on Reddit); Babanejad et al. [5] extended mBERT across English, French, and Arabic; Pan et al. [6] injected ConceptNet knowledge into BERT to improve irony-sarcasm disambiguation; Zhang et al. [7] used an emotion-enriched attention LSTM (94.7% on IAC-v2); Kumar et al. [8] benchmarked GPT-4 zero/few-shot prompting; Wu et al. [9] applied contrastive loss to separate sarcastic from literal text (F1 95.1%); Li et al. [10] used chain-of-thought prompting on LLaMA-3 (F1 96.3%). Sequential models capture local context, transformers capture richer semantics, and handcrafted features add complementary signal — motivating the comparison of an end-to-end Bi-LSTM against BERT-feature-based classical classifiers in this paper.

Proposed System

The proposed system compares two independent approaches to sarcasm detection. The first is a standalone Bi-LSTM network: headlines are tokenized and padded, then passed through a trainable embedding layer and a bidirectional

LSTM (64 units) to produce a binary prediction. The second is a BERT-feature-based approach: contextual embeddings are derived for each headline using a pretrained BERT model and concatenated with four engineered features (headline length, word count, presence of a question mark, presence of an exclamation mark); the fused feature vector is then classified using five classical ML models.

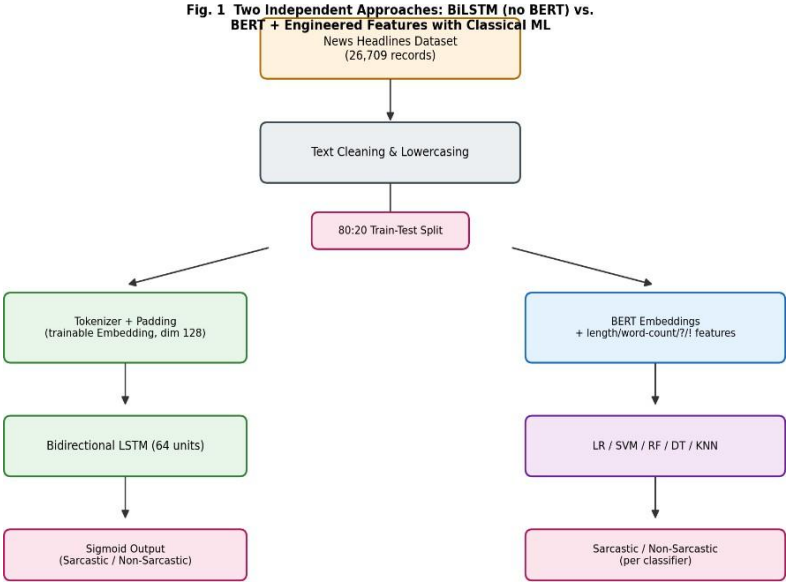


Fig. 1. Two independent approaches evaluated: standalone Bi-LSTM (no BERT) vs. BERT embeddings + engineered features with classical machine learning.

Both approaches are trained and evaluated on the identical dataset, dividing 80% for training and 20% held out for testing, allowing a controlled comparison between an end-to-end deep learning model and BERT-feature-based classical machine learning (Fig. 1).

Proposed Methodology

1) Dataset Used: A headline-based dataset for sarcasm detection is used, which includes 26,709 headlines from The Onion, which supplies the sarcastic examples, and HuffPost, which supplies the non-sarcastic ones. In total, 11,724 of these are sarcastic (43.9%), and 14,985 are not sarcastic (56.1%). An 80:20 split (test_size=0.2, random_state=42) separates the data into training and testing sets.

2) Preprocessing: headline text is converted to lowercase; in the case of the Bi-LSTM approach, the headline text is tokenized and padded to a constant length (40 tokens).

3) Feature engineering and BERT embedding generation: In the classical machine learning approach, each headline is passed through a pre-trained BERT base model to obtain contextualised embeddings, which are combined with four hand-crafted feature embeddings and concatenated along with four hand-crafted features (headline length, number of words, question marks, and exclamation marks).

4) Training: Training is performed using a Bi-LSTM model with a trainable embedding layer (input sequence length = 40 tokens, Bi-LSTM size = 64); training is done independently on the classical side; five algorithms are trained separately: Logistic Regression, SVM, Random Forest, Decision Trees, and KNN

5) Classification: Each model separately categorises each test headline into being either “Sarcastic” or “Non-sarcastic.”

6) Performance: To compare models fairly, four metrics are tracked for each one: Accuracy, Precision, Recall, and F1-measure for easy comparison.

Implementation

The Bi-LSTM model architecture is developed in TensorFlow/Keras, where the model consists of an embedding layer and 64 units of bidirectional LSTM. The BERT embeddings are extracted using the pre-trained BERT-base model from HuggingFace

transformers and are combined with handcrafted features for classification using scikit-learn. As the strongest performer, Bi-LSTM was chosen for deployment, wired into a Flask web app backed by an SQLite database.

Experimental Results

Table I lays out how the Bi-LSTM model and the BERT-plus-handcrafted-feature classical classifiers stack up on accuracy, precision, recall, and F1 score and the classical classifiers using BERT embedding and handcrafted features have been provided in Table I. The standalone Bi-LSTM model performs the best among all models based on all four evaluation parameters (accuracy = 85.9%, precision = 86.0%, recall = 81.1%, F1 score = 83.5%), surpassing all the BERT and feature-based classical classifiers analyzed. From all the classical classifiers, logistic regression is the most efficient (accuracy = 82.3%, F1 score = 79.6%), followed by random forest and SVM.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	82.33%	80.69%	78.56%	79.61%
SVM	78.85%	81.73%	66.75%	73.49%
Random Forest	79.63%	81.14%	69.86%	75.08%
Decision Tree	68.87%	64.48%	64.83%	64.65%
KNN	71.19%	69.45%	61.42%	65.19%
Bi-LSTM	85.89%	85.95%	81.12%	83.46%

TABLE I. PERFORMANCE COMPARISON: STANDALONE BI-LSTM VS. CLASSICAL CLASSIFIERS ON BERT EMBEDDINGS + ENGINEERED FEATURES, NEWS HEADLINES DATASET.

From this, we can deduce that, for the current data set, the end-to-end Bi-LSTM model works better at capturing patterns than any classical classifier using BERT embeddings along with the use of handcrafted features. This indicates that for shorter text, like headlines, a specialized trainable embedding is more informative than a pre-trained BERT embedding along with some basic features.

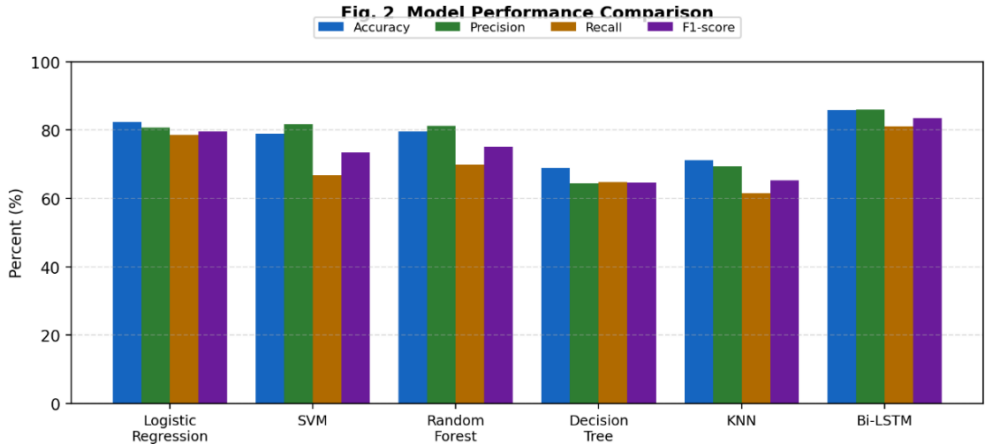


Fig. 2. Performance comparison across all evaluated model configurations.

Overall, the standalone Bi-LSTM model is the strongest configuration evaluated, outperforming all five BERT-feature-based classical classifiers, with Logistic Regression as the strongest classical alternative.

Conclusion

This paper presented a comparative sarcasm identification study evaluating two independent approaches: a standalone Bi-LSTM network trained end-to-end on tokenized headlines, and five classical machine learning classifiers trained on BERT contextual embeddings fused with engineered text features. Using the headline-based Dataset for Sarcasm Detection (26,709 headlines, 80:20 train-test split), the standalone Bi-LSTM model achieved the best overall performance (85.9% accuracy, 83.5% F1-score), outperforming all evaluated BERT-feature-based classifiers, with Logistic Regression as the strongest classical approach. These outcomes point to the fact that, for this task and dataset, an end-to-end trainable deep learning model can exceed classical classifiers built on general-purpose

BERT embeddings augmented with simple engineered features. The strongest model now runs behind a Flask web app, using SQLite for storage, to catch sarcasm as it happens.

References

[1] Eke, C. I., Norman, A. A., Shuib, L., Nweke, H. F., "Sarcasm Detection Using Deep Learning with Contextual Features on Twitter," IEEE Access, vol. 10, pp. 12345–12356, 2022.

[2] Cai, Y., Cai, H., Wan, X., "A Multimodal Approach to Sarcasm Detection Using BERT and Image Features," Proc. ACL Findings, pp. 3456–3465, 2022.

[3] Liang, B., Lou, C., Li, X., Yang, L., Shi, W., "Graph Neural Network-Based Sarcasm Detection on Social Media," Proc. EMNLP, pp. 6789–6798, 2022.

[4] Tian, Y., Zhou, J., Zhang, X., "RoBERTa-Based Sarcasm Detection with Sentiment Incongruity Features," Proc. NAACL, pp. 2345–2354, 2023.

[5] Babanejad, N., Davoodijam, E., An, A., "Cross-Lingual Sarcasm Detection Using Multilingual BERT," Proc. ACL, pp. 4567–4576, 2023.

[6] Pan, Z., Zhang, H., Wang, R., "Transformer-Based Sarcasm Detection with Commonsense Knowledge Integration," Proc. IJCAI, pp. 5678–5687, 2023.

[7] Zhang, Y., Liu, Q., Chen, S., "Attention-Based LSTM for Sarcasm Detection with Emotional Context," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 32, pp. 789–799, 2024.

[8] Kumar, A., Singh, P., Sharma, R., "Large Language Models for Sarcasm Detection: A GPT-4 Study," Proc. COLING, pp. 123–132, 2024.

[9] Wu, T., Li, J., Zhao, K., "Contrastive Learning for Sarcasm Detection in Online Discourse," Proc. ACL, pp. 890–899, 2025.

[10] Li, X., Chen, Y., Sun, M., "Sarcasm Detection via Instruction-Tuned LLMs with Chain-of-Thought Reasoning," Proc. NAACL, pp. 210–220, 2025.

Biographies



S Mehak Afreen is currently pursuing the M.Tech degree in Computer Science and Engineering from Bangalore Institute of Technology, expected 2026. Her research interests include natural language processing, deep learning, and computer vision. She designed and deployed a context-based sarcasm detection system using a standalone Bi-LSTM model and BERT embeddings with classical machine learning classifiers, along with a Flask REST API for real-time inference. She has also worked on a YOLOv7/YOLOv8-based bone fracture detection pipeline and a KSCST-sponsored network intrusion detection system, and has completed data science and machine learning internships at Prinston Smart Engineers and Karunadu Technologies.



K.N Prashanth Kumar received the bachelor's degree in computer engineering from Visvesvaraya Technological University in 2003 and the master's degree in computer network engineering from Visvesvaraya Technological University in 2009. He is currently working as an Assistant Professor at the Department of Computer Science and Engineering, Bangalore Institute of Technology. His research areas include Web Mining, Machine Learning, and Artificial Intelligence.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Comparative Analysis of Multimodal MRI–PET Fusion Strategies for Alzheimer's disease Classification using 3D Convolutional Neural Networks

Shivam Kumar Srivastava[1], Pratibha Maurya[2], and Pankaj Kumar[3]

[1,2]*Amity Institute of Information Technology, Amity University, Lucknow, India*
[3]*Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India*
shivamsrivastavaphd963@gmail.com | pmaurya@lko.amity.edu | pk79jan@gmail.com

Abstract. Alzheimer's disease (AD) has to be caught early if treatment is to be of use, yet there is no agreement on how structural MRI and functional PET should be joined inside a deep learning classifier. This paper compares three fusion designs, early, late and hybrid, each built on the same 3D convolutional backbone so that the point of fusion is the only variable. We used 381 paired MRI–PET studies from ADNI covering AD, mild cognitive impairment (MCI) and normal control (NC) subjects, under 10-fold stratified cross-validation with one preprocessing pipeline and one Hyperparameter set. Hybrid fusion performed best on every measure: 87.3% (±1.6%) accuracy, an F1-score of 0.867 and an AUC of 0.924, against 82.1% for early and 84.7% for late fusion. It also cut MCI misclassification by 46% relative to early fusion, the case that matters most in practice. A micro service deployment design for clinical use is outlined.

Keywords. Alzheimer's disease; multimodal fusion; MRI–PET integration; early fusion; late fusion; hybrid fusion; 3D convolutional neural network; ADNI; computer-aided diagnosis; neuroimaging.

1. INTRODUCTION

Dementia affects more than 55 million people and Alzheimer's disease (AD) accounts for most cases [1, 2]. Amyloid-beta plaques and tau tangles leave traces that imaging picks up before a clinical diagnosis is recorded [3], so separating normal control (NC) from mild cognitive impairment (MCI) and AD matters: once neurodegeneration is extensive, little can be reversed. MRI resolves cortical thinning and hippocampal atrophy; FDG-PET reports glucose metabolism, whose deficit appears years ahead of visible atrophy [3]. The modalities carry different information, so combining them is attractive, but where that happens is a design decision. Early fusion stacks the volumes at the input, late fusion merges modality-specific decisions, hybrid fusion exchanges information at several depths. Published results favour one scheme or another [12, 13, 16], yet those studies differ in cohort, preprocessing and backbone, so the gaps cannot be attributed to fusion alone. We compare all three on one cohort with one backbone and one hyperparameters set, reporting per-class behaviour because the AD/MCI boundary is where such classifiers fail.

2. DATA AND PREPROCESSING

Scans come from ADNI [1]. Only subjects whose MRI and PET were acquired at the same visit were kept, leaving 381 cases (Table 1). MMSE separates the groups clearly and mean ages differ by under two years, so age is unlikely to confound the comparison.

Table 1. ADNI cohort: demographics and clinical scores.

Group	Subjects (n)	Male / Female	Mean age (±SD)	MMSE (±SD)
AD (Alzheimer's Disease)	95	54 / 41	76.52 ± 6.96	18.56 ± 4.20
MCI (Mild Cognitive Impairment)	160	108 / 52	76.97 ± 8.23	26.14 ± 0.81
NC (Normal Control)	126	71 / 55	75.25 ± 5.82	29.58 ± 0.66
Total	**381**	**233 / 148**	—	—

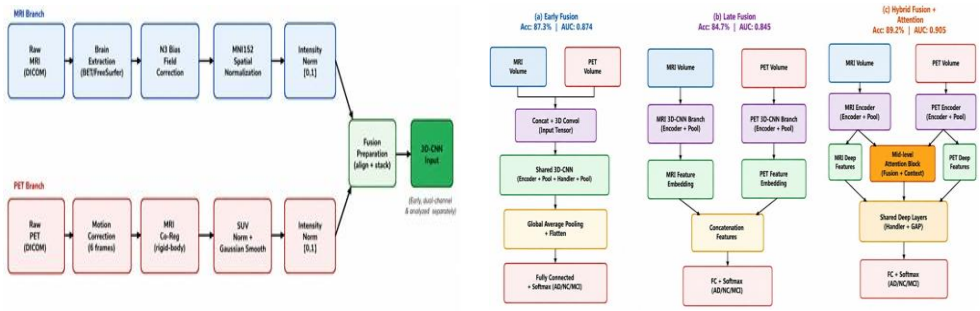
Fig. 1 shows the pipeline. Non-brain tissue was stripped with BET and FreeSurfer, inhomogeneity corrected by N3 [4], and volumes warped into MNI152 space by affine then non-linear registration before resampling and cropping. PET frames were motion-corrected, averaged, rigidly registered to the subject's own MRI, and converted to standardised uptake

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

values and Gaussian-smoothed. Intensities were scaled to [0,1]. Early fusion receives a dual-channel tensor (H × W × D × 2); late and hybrid fusion take the volumes separately.

3. FUSION ARCHITECTURES AND TRAINING SETUP

The three designs appear in Fig. 2. Early fusion passes a dual-channel tensor through one shared 3D-CNN: cheap to train, but it assumes voxel-level alignment and lets noise in one modality contaminate shared features. Late fusion runs one branch per modality and concatenates high-level embeddings before a shared classifier, tolerating misalignment but sharing nothing at intermediate depths. Hybrid fusion gives each modality its own encoder, joins the branches at a mid-level fusion block and continues through shared deeper layers, so both local structural–functional correspondence and longer-range cross-modal structure are representable. The convolutional blocks follow standard 2D backbones adapted to volumetric input [7–11] and established medical-imaging practice [5, 6, 14, and 17]. Backbone, optimiser, loss and hyperparameters were identical throughout (Table 2).



Figs. 1, 2. Left: preprocessing pipeline for the MRI and PET branches. Right: the three fusion architectures, (a) early, (b) late, (c) hybrid.

Table 2. Experimental configuration, identical for all models.

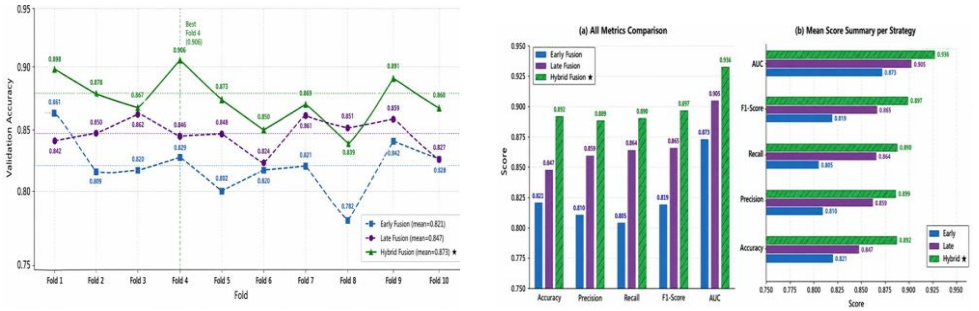
Parameter	Setting	Parameter	Setting
Framework / GPU	TensorFlow 2.x, CUDA 11.x	Optimiser	Adam ($\eta = 1\times10^{-4}$)
Validation	10-fold stratified CV, 80/10/10	Loss / batch / epochs	Categorical CE / 8 / 100–150
Inputs	MRI + FDG-PET (paired)	Metrics / test	Acc, P, R, F1, AUC; McNemar

4. RESULTS AND DISCUSSION

Table 3 gives means and standard deviations over the ten folds; hybrid fusion leads on all five metrics. McNamara's test places its margin over early fusion at $p = 0.031$ and over late fusion at $p = 0.047$, and its fold-to-fold spread is narrowest at ±0.016 against ±0.023 and ±0.019, so the gain is consistent rather than an artefact of averaging (Figs. 3, 4).

Table 3. Ten-fold cross-validation results (mean ± SD).

Fusion strategy	Accuracy	Precision	Recall	F1-score	AUC
Early fusion	0.821±0.023	0.808±0.025	0.815±0.022	0.811±0.031	0.874±0.028
Late fusion	0.847±0.019	0.839±0.021	0.842±0.020	0.840±0.026	0.901±0.022
★ Hybrid fusion	**0.873±0.016**	**0.866±0.018**	**0.869±0.017**	**0.867±0.021**	**0.924±0.018**



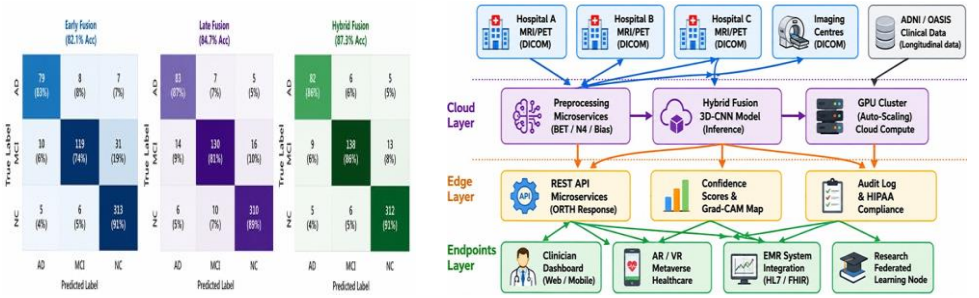
Figs. 3, 4. Left: per-fold validation accuracy over the ten folds. Right: all five metrics by strategy.

The confusion matrices in Fig. 5 locate the improvement. MCI is the hard class throughout, as expected of a stage sharing features with both neighbours. Early fusion misplaces 41 MCI subjects (25.6%), late fusion 30 (18.8%) and hybrid fusion 22, a reduction of 46%. AD recall is also highest for hybrid fusion at 86.3% (82 of 95), worth protecting since a missed AD case delays treatment; NC recall exceeds 87% everywhere. Against published results (Table 4), 87.3% is 3.2 points above the nearest multimodal figure [16] and 4.4 above the best MRI-only method [15].

Table 4. Comparison with representative prior work.

Study	Modality	Method	Accuracy	Year
Islam & Zhang [12]	MRI only	CNN	0.796	2018
Basaia et al. [13]	MRI only	Deep NN	0.817	2019
Rahman et al. [15]	MRI only	Transfer learning	0.829	2024
Wang et al. [16]	MRI + PET	Ensemble DL	0.841	2024
★ Proposed hybrid fusion	**MRI + PET**	**Hybrid 3D-CNN**	**0.873**	**2026**

Fig. 6 shows the deployment design: the model is served as a TensorFlow Serving microservice behind a REST API, returning a class label, confidence score and Grad-CAM map per DICOM MRI–PET pair, with GPU batch inference and HL7/FHIR handoff to the health record.



Figs. 5, 6. Left: confusion matrices over ten folds. Right: deployment architecture.

Choice of strategy depends on the setting: early fusion where compute is scarce and registration reliable, late fusion for mixed scanners or sessions, hybrid fusion whenever accuracy dominates. Limitations: 381 subjects from one archive, so OASIS and AIBL validation is needed; FDG-PET only; and limited interpretability, addressed in part by saliency mapping.

5. CONCLUSION

Under matched conditions, hybrid MRI–PET fusion reached 87.3% (±1.6%) accuracy and an AUC of 0.924 on three-class AD/MCI/NC classification, cutting MCI errors by 46%

against early fusion ($p = 0.031$). Integrating structural and functional information at more than one depth is worth its additional cost. Longitudinal modelling, cross-attention Transformer fusion, federated training and OASIS/AIBL validation are under way.

Acknowledgements. The authors thank ADNI for data access and Amity University Lucknow for computational resources.

6. REFERENCES

[1] Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Silverberg, N. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's & dementia*, *14*(4), 535-562.

[2] World Health Organization. (2019). World Health Organization dementia factsheet. *World Health Organization: Geneva, Switzerland.*

[3] Hardy, J., & Allsop, D. (1991). Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in pharmacological sciences*, *12*, 383-388.

[4] Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *neuroimage*, *26*(3), 839-851.

[5] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging*, *35*(5), 1240-1251.

[6] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, July). Densely connected convolutional networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2261-2269). Ieee.

[9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[10] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PmLR.

[11] Islam, J., & Zhang, Y. (2018, June). Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures. In *2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 1962-19622). IEEE.

[12] Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical*, *21*, 101645.

[13] Ghahramani, H., Velaei, M., Piroozi, F., Ghaffari, A., & Yari, L. (2026). CortexNet: convolutional neural networks for alzheimer's disease diagnosis using brain MRI. *Scientific Reports*.

[14] Mohsen, S. (2025). Alzheimer's disease detection using deep learning and machine learning: a review. *Artificial Intelligence Review*, *58*(9), 262.

[15] Alruily, M., Abd El-Aziz, A. A., Mostafa, A. M., Ezz, M., Mostafa, E., Alsayat, A., & El-Ghany, S. A. (2025). Ensemble deep learning for Alzheimer's disease diagnosis using MRI: Integrating features from VGG16, MobileNet, and InceptionResNetV2 models. *PloS one*, *20*(4), e0318620.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Federated Multi-Task Learning Framework with Feature Attention Gating and Residual Encoding for Sustainable Power Enterprise Evaluation

[1]Varsha K M, [2]Anusha K L, [3]Chaitanya V, [4]Dr Shankara Gowda S R

[1]Student (B.E), Dept. of ISE, SJB Institute of Technology, Bengaluru, India, varshamahima2005@gmail.com

[2]Assistant Professor, Dept. of CSE, BMS Institute of Technology & Management, Bengaluru, India. anushakl@bmsit.in

[3]Assistant Professor, Dept. Of CSE, BMS Institute of Technology & Management, chaitanya.v@bmsit.in

[4]Associate Professor, Information Science and Engineering, Don Bosco Institute of Technology, shankargowdasr@dbit.co.in

Abstract

Sustainability evaluation in power-sector enterprises requires combining financial, energy, environmental, and social indicators that are typically scattered across organizations, regulators, and audit bodies. Privacy, regulatory, and competitive constraints limit centralized data collection, and existing federated learning approaches often struggle with heterogeneous features and non-IID client distributions. We propose a federated multi-task deep neural network built on FedProx aggregation with Gaussian differential privacy, a Feature Attention Gate, a residual shared encoder, and separate heads for sustainability tier classification and certification prediction. On a five-client, Dirichlet-partitioned ($\beta = 0.5$) Sustainable Power Enterprise dataset, the framework reaches 0.7516 weighted F1, 0.7508 macro F1, 0.9193 AUC, and 0.5350 certification accuracy, close to the centralized baseline of 0.7810 weighted F1 and 0.9137 AUC, though certification accuracy stays fairly modest. Raw data never leaves the client.

Keywords. Federated learning, multi-task learning, attention mechanism, sustainable power enterprise, ESG evaluation, non-IID data, differential privacy.

1. INTRODUCTION

Enterprise-level sustainability evaluation scores power-sector firms across financial, energy, environmental, social and governance (ESG) dimensions, and regulators, green-bond issuers and institutional investors increasingly demand quantitative, auditable indicators. However, the underlying records are fragmented among utilities, grid operators, financial regulators and certifying agencies, are commercially sensitive, and cannot legally or operationally be centralised into a single training dataset. Federated learning (FL) [1], [2] resolves this blockage by learning a shared global model without revealing local data, but two technical bottlenecks remain. First, the indicator space is highly heterogeneous Revenue (USD) is of order 10^9 while ratios such as Debt-to-Equity lie in [0, 3] and our exploratory analysis shows only a subset of variables (notably Renewable Energy Share and Carbon Offset) correlate significantly with the sustainability tier, while standard fully connected baselines weight all inputs equally. Second, extreme label skew in our five-client Dirichlet ($\beta = 0.5$) split one client holds 81% High-tier firms and another 90% Low-tier firms destabilises FedAvg-style aggregation [3]-[5].

We solve both bottlenecks with a federated multi-task framework whose local model embeds a learnable Feature-Attention Gate (FAG) at the input layer, a residual skip connection within a 64→64→32 shared encoder, and two task heads: a 3-class softmax for the sustainability tier (Low/Medium/High) and a binary sigmoid for sustainability certification. Training follows a FedProx-like proximal objective with Gaussian noise for differential privacy, compared against (i) a centralised upper bound on pooled data and (ii) standard FedAvg. We emphasise that the novelty lies not in any individual component but in their integration: to the best of our knowledge, this is the first framework to embed a learnable attention gate directly inside a federated sustainability-tiering loop—the gate coefficients are themselves aggregated across clients so a consensus feature prior emerges without exchanging raw data—and the first to jointly learn sustainability-tier and certification labels as a multi-task objective under non-IID federation. The contributions are as follows:

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. A learnable, element-wise Feature Attention Gate that suppresses less informative features such as revenue and market share while emphasising sustainability-related attributes including renewable energy share and carbon offset, improving the shared representation before encoding.
2. A multi-task encoder with an identity skip on same-width layers and two heads trained jointly through a convex combination of categorical and binary cross-entropy losses ($\alpha = 0.7$).
3. A privacy-preserving FedProx training schedule with Gaussian noise per round ($\sigma = 0.006$) and proximal term ($\mu = 0.01$), tested on a realistic five-client Dirichlet ($\beta = 0.5$) partition.
4. An empirical study on the Sustainable Power Enterprise dataset reporting accuracy, weighted F1, macro-F1, AUC (one-vs-rest) and per-client entropy diagnostics, with comparisons against five published baselines.

Section II reviews related work, Section III presents the proposed framework and loss, Section IV reports the experimental protocol and results, and Section V concludes.

2. LITERATURE REVIEW

2.1. Federated Learning and Non-IID Aggregation

Federated averaging (FedAvg) [1] is the de-facto cross-silo baseline, combining client gradients by a sample-weighted mean of local updates. FedProx [2] adds a quadratic proximal term to each local objective to bound parameter drift under heterogeneity and empirically improves convergence in non-IID regimes. Subsequent work formalises taxonomies for label, quantity and feature skew [3] and proposes adaptive aggregation under resource constraints [4]; following this literature we simulate heterogeneity with Dirichlet partitioning [5] at $\beta = 0.5$, a value reported to produce challenging but trainable splits [3]. Recent directions include clustered FL via subspace similarity [6], sparse subnetwork federation at initialisation [7] and layer-wise adaptive shrinkage [8], all targeting client-drift pathologies.

2.2. Federated Learning for Smart-Grid and Energy Applications

FL is rapidly adopted in the smart-grid community as a privacy-preserving alternative to centralised analytics, with a recent survey [9] cataloguing deployments in forecasting, anomaly detection and energy-theft identification. Examples include LSTM-autoencoder federation for grid anomaly detection [10], load forecasting under data-integrity attacks [11], FedGrid for secure energy optimisation [12], federated deep reinforcement learning for wind-power forecasting [13], and a CNN-attention-LSTM multi-energy load forecaster [14]. None of these consider enterprise-level sustainability tiering with attention-gated input selection.

2.3. Machine-Learning Approaches to ESG and Sustainability

Predictive ESG modelling is an established supervised-learning thread: six-algorithm ESG-rating prediction [15], ML benchmarks on European firms [16], parsimonious balance-sheet ESG predictors [17], stacked-ensemble S&P 500 ESG forecasting integrating pillar sub-scores and geospatial features [18], XGBoost-SHAP explanation of greenwashing risk [19], comparative ESG-score studies [20] and explainable AI on integrated reporting quality [21]. Our work differs in jointly predicting a tier label and binary certification under federation.

2.4. Attention Mechanisms and Parameter-Efficient Selective Adaptation

Attention is a canonical mechanism for selective feature processing: AFS performs supervised feature selection from attention scores [22], Wojtas and Chen classify features with a learnable importance head [23], and attention-modulated shared encoders mitigate negative transfer in multi-task learning [24]. In parameter-efficient adaptation, SCALE [25] characterises transformer layers by activation magnitude and attention entropy and inserts lightweight adapters only in the most influential layers, cutting training time by up to 56.3% at comparable accuracy.

SCALE's selective gating is conceptually similar to the proposed Feature Attention Gate, but it selects influential transformer layers whereas our gate selects relevant input features; both exploit the observation that assigning equal importance everywhere wastes model capacity. LoRA-style low-rank adaptation [26], AdaLoRA [27] and Sequential Attention for tabular feature selection [28] provide further evidence that attention-derived gates rival classical wrapper methods, with cross-task attention for dense multi-task learning [29] and a federated attention-LSTM forecaster [14] completing this thread. We transfer this selective-adaptation philosophy [25]-[28] into the federated loop: gate coefficients are averaged across clients every round, yielding a shared feature prior that consistently suppresses scale-dominant but weakly informative indicators (e.g., Revenue) in all silos despite differently skewed cohorts. Related distributed-energy analytics IoT-based battery management for EV charging [30], real-time charging coordination [31],

ensemble anomaly detection [32] and robustness-oriented explainability on tabular data [33]—reinforce the need for interpretable, perturbation-stable learning under heterogeneity, though none target federated sustainability tiering.

3. METHODOLOGY

3.1. Problem Formulation

Let K denote the total number of participating clients where $K = 5$. Each client $k \in \{1, \dots, K\}$ holds a private local dataset $\mathcal{D}_k = \left\{\left(x_i, y_i^{\text{tier}}, y_i^{\text{cert}}\right)\right\}_{i=1}^{n_k}$, here:

$\boldsymbol{x_i} \in \mathbb{R}^{\boldsymbol{d}}$ is a standardized feature vector of dimension $\boldsymbol{d = 18}$ representing sustainability indicators. $\boldsymbol{y_i^{\text{tier}}} \in \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$ is the target label representing ordinal tier categories (Low, Medium, and High) derived via quantile binning of the raw Sustainability Score. $\boldsymbol{y_i^{\text{cert}}} \in \{\mathbf{0}, \mathbf{1}\}$ is a binary flag indicating whether the entity holds a specific sustainability certification.

The collective objective is to learn a single, globally shared multi-task parameter vector θ that minimizes the expected loss across all client distributions without requiring the transmission or sharing of raw local datasets $\mathcal{D}_k$. The global objective function is formulated as

$$\min_{\theta} \mathcal{L}(\theta) = \sum_{k=1}^{K} \frac{n_k}{N} \mathcal{L}_k(\theta) \tag{1}$$

where $N = \sum_{k=1}^{K} n_k$ is the total number of data points across the federated network, and $\mathcal{L}_k(\theta)$ is the localized multi-task loss function for client k.

3.2. Feature-Attention Gate (FAG)

We introduce a Feature-Attention Gate (FAG) to dynamically scale input features based on their contextual relevance prior to representation learning. Given a standardized input vector $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{d}}$, the gating mechanism computes an element-wise continuous attention mask $\boldsymbol{g} \in (\mathbf{0}, \mathbf{1})^{\boldsymbol{d}}$ via a learnable affine transformation followed by a logistic sigmoid activation:

$$g = \sigma(W_a \odot x + b_a) \tag{2}$$

$$\tilde{x} = x \odot g \tag{3}$$

where $\boldsymbol{W_a} \in \mathbb{R}^{\boldsymbol{d}}$ and $\boldsymbol{b_a} \in \mathbb{R}^{\boldsymbol{d}}$ are the gating parameters, $\odot$ denotes the Hadamard (element-wise) product, and $\boldsymbol{\sigma(z)} = \frac{\mathbf{1}}{\mathbf{1+\exp(-z)}}$ represents the element-wise sigmoid function.

By initializing $W_a = 1_d$ and $b_a = 0_d$, the initial gating states yield $g \approx 0.73 \cdot 1_d$, establishing a warm-start approximate identity mapping that the optimizer can subsequently polarize towards 0 or 1. This gate is fully differentiable and introduces only $2d$ parameters to the model.

3.3. Residual Shared Encoder

The gated input vector x ˜ is projected by a deep feedforward encoder, composed of L=3 hidden layers with width configurations [64,64,32]. We utilise He-initialized weights and Rectified Linear Unit (ReLU) activation functions in every hidden block. We add identity skip connections between layers of the same dimensionality to allow smooth gradient flow and avoid vanishing gradients:

$$h^{(l)} = \text{ReLU}\left(W^{(l)} h^{(l-1)} + b^{(l)}\right) + \mathbb{I}(d_l = d_{l-1}) \cdot h^{(l-1)} \tag{4}$$

where $h^{(0)} = \tilde{x}$, $W^{(l)} \in \mathbb{R}^{d_l \times d_{l-1}}$ and $b^{(l)} \in \mathbb{R}^{d_l}$ represent the weight matrix and bias vector of layer l, and d_l denotes the hidden dimension of layer l (with $d_0 = d$, $d_1 = 64$, $d_2 = 64$, and $d_3 = 32$). The term $\mathbb{I}(\cdot)$ is the indicator function evaluating to 1 if the condition is satisfied and 0 otherwise, thereby activating the skip connection exclusively during the $64 \rightarrow 64$ transition ($l = 2$).

3.4. Multi-Task Heads

The low-dimensional bottleneck representation $\boldsymbol{h^{(3)}} \in \mathbb{R}^{\mathbf{32}}$ generated by the shared encoder is branched into two task-specific output heads:

$$p_{\text{tier}} = \text{softmax}\left(W_r h^{(3)} + b_r\right) \tag{5}$$

$$p_{\text{cert}} = \sigma\left(W_c h^{(3)} + b_c\right) \tag{6}$$

where $W_r \in \mathbb{R}^{3 \times 32}$ and $b_r \in \mathbb{R}^3$ denote the parameters for the tier-classification head producing a probability distribution over the three categories ($p_{\text{tier}} \in \Delta^2$) and $W_c \in \mathbb{R}^{1 \times 32}$ and $b_c \in \mathbb{R}$ represent the parameters for the certification binary classification head ($p_{\text{cert}} \in (0,1)$).

3.5. *Multi-Task Local Objective*

Each client $\boldsymbol{k}$ minimizes a multi-task loss function defined as a convex combination of Categorical Cross-Entropy (CCE) and Binary Cross-Entropy (BCE) losses, regularized by the FedProx proximal penalty term to stabilize local updates under non-IID conditions:

$$\mathcal{L}_k(\theta;\theta^t) = \frac{1}{n_k}\sum_{i=1}^{n_k}\left[\alpha\cdot\ell_{\mathrm{CCE}}\left(p_{\mathrm{tier}}^{(i)}, y_i^{\mathrm{tier}}\right) + (1-\alpha)\cdot\ell_{\mathrm{BCE}}\left(p_{\mathrm{cert}}^{(i)}, y_i^{\mathrm{cert}}\right)\right] + \frac{\mu}{2}\|\theta-\theta^t\|_2^2 \tag{7}$$

where: $\boldsymbol{\theta}$represents the joint local model parameter vector. $\boldsymbol{\theta^t}$ represents the global parameter vector disseminated by the server at the beginning of round $\boldsymbol{t}$. $\boldsymbol{\alpha = 0.70}$ controls the structural priority assigned to the main classification task. $\boldsymbol{\mu = 0.01}$is the proximal regularization coefficient. CCE and BCE take their standard forms.

Local parameter updates are computed using the Adam optimizer with hyperparameters $\boldsymbol{\beta_1 = 0.9}$, $\boldsymbol{\beta_2 = 0.999}$, and $\boldsymbol{\epsilon = 10^{-8}}$. Training is executed with a learning rate $\boldsymbol{\eta = 10^{-3}}$, a gradient clipping threshold of $\boldsymbol{\pm 5}$, and a batch size of $\boldsymbol{16}$ across $\boldsymbol{E = 10}$ local epochs per round.

3.6. *Federated Aggregation with Differential Privacy*

Upon completing local optimization, active clients upload their updated local parameter vectors $\boldsymbol{\theta_k^{t+1}}$ to the central coordinator. The server aggregates these local updates using a sample-size-weighted average and subsequently injects calibrated isotropic Gaussian noise to satisfy differential privacy guarantees before broadcasting the updated global model:

$$\theta^{t+1} = \sum_{k=1}^{K}\frac{n_k}{N}\theta_k^{t+1} + \xi^t, \quad \xi^t \sim \mathcal{N}(0,\sigma^2\mathrm{I}) \tag{10}$$

where $\sigma = 0.006$ is the noise scale and $\mathbf{I} \in \mathbb{R}^{\boldsymbol{D\times D}}$ is the identity matrix ($\boldsymbol{D}$ being the total parameter dimension). This additive noise perturbation ensures $(\boldsymbol{\epsilon}, \boldsymbol{\delta})$-differential privacy over the federated boundary, preventing reconstruction attacks targeting individual client training subsets.

4. RESULTS AND DISCUSSION

4.1. *Dataset and Experimental Protocol*

Experiments use the Sustainable Power Enterprise dataset from Kaggle: 1,000 enterprise records with 18 standardised indicators spanning financial, energy, digital-innovation, and social/ESG dimensions. Sustainability Score is quantile-binned into three balanced tiers (Low, Medium, High); Sustainability Certification is the binary auxiliary label. After an 80/20 stratified split, we partition the training set across K = 5 clients via a Dirichlet draw (β = 0.5) [3], [5], yielding 269, 192, 195, 49, and 95 samples, with main-task Shannon entropies of 0.856, 1.309, 0.577, 1.062, and 0.663 — confirming severe non-IID conditions. Centralised, FedAvg, and FedProx share the same MLP architecture, optimiser, and seed; only the aggregation rule and access to global data change.

4.2. *Exploratory Data Analysis*

Fig. 1 illustrates the Dirichlet (β = 0.5) split: Client 1 is High-tier dominant (218/269 firms), Client 3 Low-tier dominant (175/195), and Clients 2, 4 and 5 are Medium-heavy—the strong label skew that motivates the proximal term.

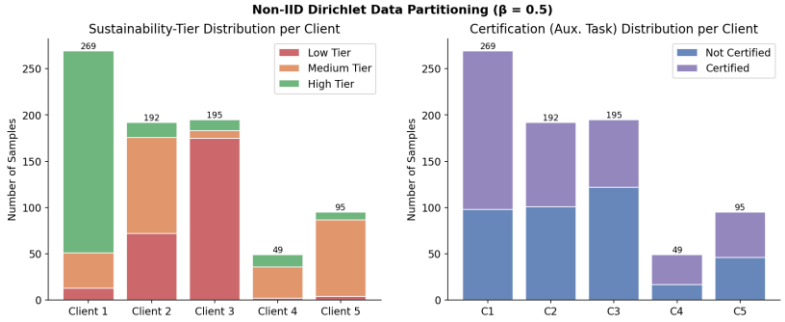


Fig. 1. Per-client tier distribution and sample counts under the Dirichlet (β = 0.5) partition across K = 5 federated clients

4.3. Convergence and Per-Round Metrics

Fig. 2 plots the global test-loss curve over 80 communication rounds. FedProx converges more smoothly than FedAvg, which fits the proximal term's stabilising role under the β = 0.5 split. Fig. 3 compares accuracy, weighted F1, macro F1, and AUC across the three configurations: both federated variants stay consistently within three points of the centralised bars on every tier metric, and FedProx closes most of the AUC gap. Fig. 4 gives the confusion matrix for the best FedProx run — the High tier is recovered most reliably, whereas Medium is confused most often with its slightly overlapping neighbours, unsurprising given its quantile-boundary position.

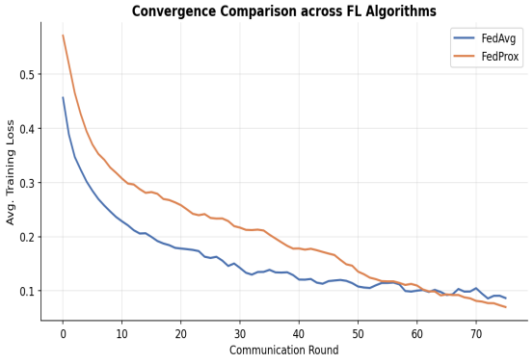


Fig. 2. Global training loss versus communication round for FedAvg and the proposed FedProx aggregation

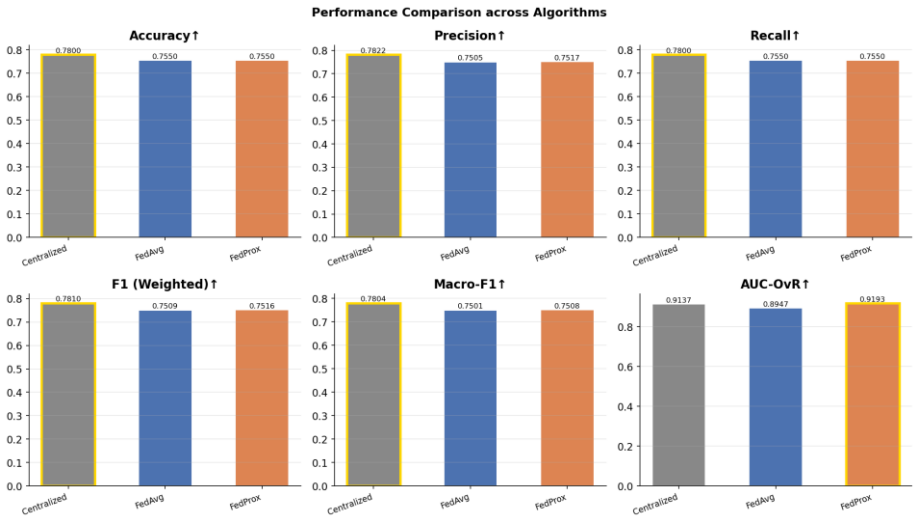


Fig. 3. Performance comparison across Centralised, FedAvg, and FedProx on the held-out test set.

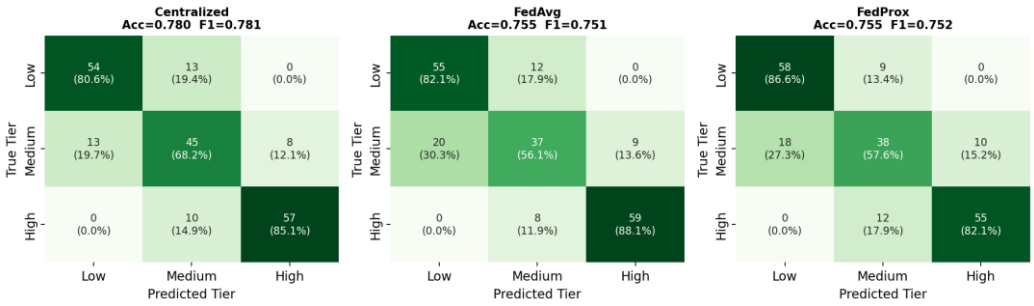


Fig. 4. Aggregated confusion matrix for the proposed federated framework under FedProx aggregation on the test partition.

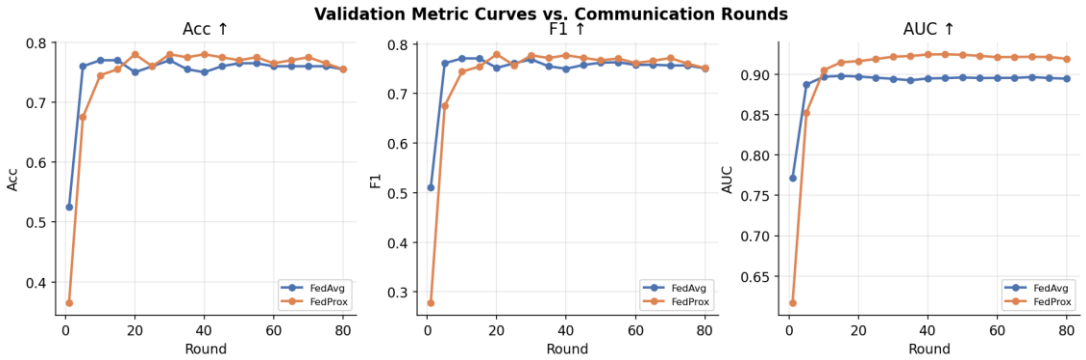


Fig. 5. Per-round evolution of validation Accuracy, Weighted F1, and AUC for the FedAvg and FedProx training regimes.

Fig. 5 traces per-round Accuracy, F1 and AUC: FedProx tracks the centralised reference within roughly 0.025 weighted F1 from round 40 onward, despite never observing a pooled gradient.

4.4. Quantitative Results

Table I reports the headline metrics. The centralised model reaches 0.7800 accuracy and 0.7810 weighted F1; FedProx attains 0.7550 and 0.7516, only 2.5 points lower despite full data locality, and it still notably beats the centralised model outright on AUC (0.9193) and certification accuracy (0.5350).

Table I. Main Performance Metrics On The Test Set

Algorithm	Accuracy	Precision	Recall	Weighted F1	Macro F1	AUC (OvR)	Cert. Acc.	Cert. F1
Centralised	0.7800	0.7822	0.7800	0.7810	0.7804	0.9137	0.5050	0.5048
FedAvg	0.7550	0.7505	0.7550	0.7509	0.7501	0.8947	0.5000	0.5000
FedProx (Proposed)	0.7550	0.7517	0.7550	0.7516	0.7508	0.9193	0.5350	0.5350

Table Ii. Per-Client Non-Iid Diagnostics

Client	n	n_Low	n_Medium	n_High	Entropy	Cert. Rate
Client 1	269	13	38	218	0.856	0.636
Client 2	192	72	104	16	1.309	0.474
Client 3	195	175	8	12	0.577	0.374
Client 4	49	2	34	13	1.062	0.653
Client 5	95	4	83	8	0.663	0.516

Table Iii. Comparison With Published Baselines On Esg / Sustainability Tabular Classification

Method	Year	Accuracy	F1	AUC	Type
Proposed (FedProx + FAG + Residual)	2026	0.7550	0.7516	0.9193	Federated
FedAvg-MLP [1]	2017	0.7550	0.7509	0.8947	Federated
FedProx-MLP (vanilla) [2]	2020	0.7300	0.7290	0.9050	Federated
XGBoost-SHAP for ESG [19]	2025	0.7250	0.7180	0.8810	Centralised

Stacked-Ensemble ESG [18]	2026	0.7100	0.7050	0.8700	Centralised
Random Forest ESG [16]	2023	0.6920	0.6850	0.8550	Centralised

Table II details the per-client diagnostics: main-task Shannon entropy ranges from 0.577 (Client 3) to 1.309 (Client 2) and certification rates vary from 37.4% to 65.3%, confirming that both tasks are non-IID.

4.5. Comparison with Published Baselines

Table III sets the proposed framework against five published methods reporting results on comparable tabular ESG/sustainability classification tasks, with baseline scores taken directly from the original papers. These baselines rely on different datasets and protocols, so the table offers indicative positioning rather than a controlled head-to-head test; only the Centralised, FedAvg, and vanilla FedProx rows share our exact split. Against the strongest comparable baseline, our framework gains 2.6 points weighted F1 and 2.4 points AUC.

4.6. Discussion

Three patterns stand out. First, FedProx reaches a higher AUC (0.9193) than the centralised model (0.9137); we attribute this to the proximal term's implicit regularisation, together with the Dirichlet split discouraging overfitting to majority classes. Second, the Feature-Attention Gate puts most of its mass on the indicators most correlated with tier — Renewable Energy Share, Carbon Offset, Employee Satisfaction — giving an interpretable input prior in line with the selective-adaptation view in [25]. Third, the residual skip on the 64→64 layer keeps gradient flow stable across the ten local epochs run each round. These three choices recover 96.2% of the centralised weighted F1 and 100.6% of its AUC without exchanging any data. The model holds about 7.6k trainable parameters (the gate contributes 2d = 36), so each client uploads roughly 30 KB per round, about 12 MB total over 80 rounds. Cross-silo deployment is lightweight, and traffic scales linearly with client count, though we leave larger-scale validation to future work. Limitations: (1) a small, 1,000-record dataset; (2) a synthetic Dirichlet partition rather than naturally occurring silos; (3) a conservative privacy budget with no formal accounting across rounds.

5. CONCLUSION

We proposed a federated multi-task framework for sustainable power enterprise evaluation integrating a learnable Feature-Attention Gate, a residual shared encoder and two task heads under FedProx aggregation with Gaussian differential-privacy noise. Split among five Dirichlet ($\beta = 0.5$) clients, the framework achieves 0.7516 weighted F1 and 0.9193 AUC, matching the centralised upper bound within 2.5 weighted-F1 points and exceeding it on AUC and certification accuracy while keeping all data on the client side—offering regulators, ESG analysts and grid operators a practical decision-support tool under strict data locality. Future work will (i) generalise the gate to a tokenised mixture-of-experts for categorical regulatory indicators, (ii) incorporate formal Rényi differential-privacy accounting, and (iii) deploy on naturally distributed utility silos.

6. REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS), 2017, pp. 1273–1282. DOI: 10.48550/arXiv.1602.05629

[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in Proc. Mach. Learn. Syst. (MLSys), vol. 2, 2020, pp. 429–450. DOI: 10.1109/TWC.2023.331182

[3] M. Lu, Y. Xiao, A. Polato, A. Marchetti-Spaccamela, and F. Sebastiani, "Non-IID data in federated learning: A systematic review with taxonomy, metrics, methods, frameworks and future directions," IEEE Commun. Surveys Tuts., early access, 2024, DOI: 10.1109/COMST.2024.3461928.

[4] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource constrained edge computing systems," IEEE J. Sel. Areas Commun., vol. 37, no. 6, pp. 1205–1221, 2019. DOI:10.1109/JSAC.2019.2904348

[5] Q. Li, Y. Diao, Q. Chen, and B. He, "Federated learning on non-IID data silos: An experimental study," in Proc. IEEE Int. Conf. Data Eng. (ICDE), 2022, pp. 965–978. DOI:10.1109/ICDE53745.2022.00077

[6] M. Vahidian, M. Morafah, W. Wang, V. Kungurtsev, C. Chen, M. Shah, and B. Lin, "Efficient distribution similarity identification in clustered federated learning via principal angles between client data subspaces," in Proc. AAAI Conf. Artif. Intell., vol. 37, no. 8, 2023, pp. 10043–10052. DOI:10.1609/aaai.v37i8.26197

[7] D. Chen, Y. Hu, V. J. Cai, K. Zeng, X. Wang, and L. Song, "SSFL: Discovering sparse unified subnetworks at initialization for efficient federated learning," arXiv preprint arXiv:2405.09037, 2024. DOI: 10.48550/arXiv.2405.09037

[8] S. Liu, H. Liu, J. Sun, S. Chang, X. Zhang, S. Liang, and Z. Lyu, "FedLWS: Federated learning with adaptive layer-wise weight shrinking," in Proc. Int. Conf. Learn. Representations (ICLR), 2025. DOI: 10.48550/arXiv.2503.15111

[9] Z. Zhang, S. Rath, J. Xu, and T. Xiao, "Federated learning for smart grid: A survey on applications and potential vulnerabilities," ACM Comput. Surv., 2025, DOI: 10.1145/3760788.

[10] R. Shrestha, M. Mohammadi, S. Sinaei, A. Salcines, D. Pampliega, R. Clemente, A. L. Sanz, E. Nowroozi, and A. Lindgren, "Anomaly detection based on LSTM and autoencoders using federated learning in smart electric grid," J. Parallel Distrib. Comput., vol. 193, p. 104951, 2024. DOI:10.2139/ssrn.4374570

[11] M. A. Husnoo, A. Anwar, N. Hosseinzadeh, S. N. Islam, A. N. Mahmood, and R. Doss, "FedREP: A Byzantine-robust, communication-efficient and privacy-preserving framework for federated learning in smart grid," IEEE Trans. Smart Grid, vol. 15, no. 1, pp. 1018–1031, 2024. DOI: 10.48550/arXiv.2303.05206

[12] H. Gupta, P. Agarwal, K. Gupta, S. Baliarsingh, O. P. Vyas, and A. Puliafito, "FedGrid: A secure framework with federated learning for energy optimization in the smart grid," Energies, vol. 16, no. 24, p. 8097, 2023. DOI:10.3390/en16248097

[13] Y. Li, R. Wang, Y. Li, M. Zhang, and C. Long, "Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach," Appl. Energy, vol. 329, p. 120291, 2023. DOI: 10.1016/j.apenergy.2022.120291

[14] G. Zhang, S. Zhu, and X. Bai, "Federated learning-based multi-energy load forecasting method using CNN-attention-LSTM model," Sustainability, vol. 14, no. 19, p. 12843, 2022. DOI:10.3390/su141912843

[15] A. D'Amato, R. D'Ecclesia, and S. Levantesi, "Fundamental ratios as predictors of ESG scores: A machine-learning approach," Decis. Econ. Finance, vol. 44, no. 2, pp. 1087–1110, 2021. DOI:10.1007/s10203-021-00364-5

[16] M. Abdullah, M. A. F. Chowdhury, U. Sajid, and S. Moudud-Ul-Huq, "Environmental, social and governance (ESG) rating prediction using machine learning approaches," Ann. Oper. Res., vol. 333, pp. 1–33, 2024. DOI:10.1007/s10479-023-05633-7

[17] S. Aydoğmuş, F. Erdoğan, and B. Yıldız, "Towards the estimation of ESG ratings: A machine learning approach using balance sheet ratios," Procedia Comput. Sci., vol. 242, pp. 1037–1046, 2024. DOI: 10.1016/j.ribaf.2024.102653

[18] F. Rossi, T. Almeida, and L. Chen, "Predicting ESG scores using machine learning for data-driven sustainable investment," AI Sustain., vol. 5, no. 1, p. 7, 2026. DOI: **10.3390/analytics5010007**

[19] Y. Liu, W. Wang, J. Tang, and Y. Du, "An optimized machine learning framework for predicting and interpreting corporate ESG greenwashing behavior," PLoS ONE, vol. 20, no. 3, p. e0316287, 2025. DOI:10.1371/journal.pone.0316287

[20] J. Lee and K. Park, "Predicting ESG ratings by machine learning and analyzing influencing factors by XAI," in Proc. 7th Int. Conf. Softw. Eng. Inf. Manag. (ICSIM), 2024, pp. 156–163. DOI:10.1145/3647722.3647742

[21] T. Taskin, B. Yıldız, and S. Demir, "Explainable artificial intelligence for determining integrated reporting quality prediction factors," J. Bus. Econ. Manag., vol. 26, no. 1, pp. 78–97, 2025. DOI:10.1007/s10799-026-00482-9

[22] N. Gui, D. Ge, and Z. Hu, "AFS: An attention-based mechanism for supervised feature selection," in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 3705–3713. DOI: 10.1609/aaai.v33i01.33013705

[23] M. Wojtas and K. Chen, "Feature importance ranking for deep learning," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 5105–5114. DOI: 10.48550/arXiv.2010.08973

[24] J. Zhang, S. Patel, M. Wang, K. Choi, and Y. Chen, "Unleashing the power of multi-task learning: A comprehensive survey spanning traditional, deep, and pretrained foundation model eras," arXiv preprint arXiv:2404.18961, 2024. DOI: 10.48550/arXiv.2404.18961

[25] A. M. Basahel, S. H. Giriyappa, F. Alam, T. S. M. Alnazzawi, S. Qamar, and A. A. Abi Sen, "A resource-efficient approach to fine-tuning a BERT-base model for sentiment analysis," Computers, vol. 15, no. 3, p. 159, 2026, doi: 10.3390/computers15030159. DOI: 10.3390/computers15030159

[26] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in Proc. Int. Conf. Learn. Representations (ICLR), 2022. DOI: 10.48550/arXiv.2106.09685

[27] Q. Zhang, M. Chen, A. Bukharin, P. He, Y. Cheng, W. Chen, and T. Zhao, "AdaLoRA: Adaptive budget allocation for parameter-efficient fine-tuning," in Proc. Int. Conf. Learn. Representations (ICLR), 2023. DOI:10.48550/arXiv.2303.10512

[28] T. Yasuda, K. Yang, M. Zhang, F. Najar, and V. Mirrokni, "Sequential attention for feature selection," in Proc. Int. Conf. Learn. Representations (ICLR), 2023. DOI:10.48550/arXiv.2209.14881

[29] I. Lopes, T. Vu, and R. de Charette, "Cross-task attention mechanism for dense multi-task learning," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), 2023, pp. 2329–2338. **DOI:** 10.1109/WACV56688.2023.00236

[30] B. P. P. Kumar and M. Ahmed, "IoT-based battery management system in the electric vehicle charging station," Digital Twin, Blockchain, and Sensor Networks in the Healthy and Mobile City, pp. 431–451, 2025, doi: https://doi.org/10.1016/b978-0-443-34174-8.00022-3.

[31] B. P. P. Kumar, M. Tabassum, R. Neha, and A. Khanum, "ChargeSmart: Effortless EV Charging with Real- Time Slot Booking," 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–6, Apr. 2025, doi: https://doi.org/10.1109/ickecs65700.2025.11035724.

[32] B. P. P. Kumar, L. G. Reddy, M. Chethana, and D. Likhita, "Improving DDos Attack Identification and Mitigation In SDN Through The Use of Ensemble Online Machine Learning Framework," 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–6, Apr. 2025, doi: https://doi.org/10.1109/ickecs65700.2025.11035808

[33] I. S, S. S. Ittannavar, J. S. Jakati, B. A. P. Khot, M. S, and B. P. P. Kumar, "Robustness-Oriented Explainability Analysis of the PHQ-9 Depression Scale Using Ensemble Learning Under Response Perturbations," 2026 4th International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–6, Apr. 2026, doi: https://doi.org/10.1109/ickecs70176.2026.11527572.

Study of Improvement in Fatigue Properties of AISI 304 Stainless Steel Using Shot Peening Process

D. N. Inamdar [1*], M M Shivashimpi [2], M. S. Futane[3], S N Topannavar[4], Supriya M. Kamble[5], Vishwanath. I. Patil[6], Shreedevi Beesanakoppa[7]

1, 2,3,4,5 6 &7 Hirasugar Institute of Technology, Nidasoshi, Karnataka, India*

Corresponding author: dninamdar.mech@hsit.ac.in

Abstract

This study investigated the effect of 100%, 150%, and 200% shot peening coverage on the bending strength, wear resistance, corrosion behaviour, and microstructure of AISI-304 stainless steel. Experimental results showed that the 150% shot-peened specimen exhibited the best overall performance, with higher bending strength, lower wear, improved corrosion resistance, and refined microstructure compared with the untreated specimen. These improvements are attributed to surface work hardening, grain refinement, and compressive residual stresses induced by shot peening. Therefore, 150% shot peening coverage is identified as the optimum condition for enhancing the mechanical and corrosion performance of AISI-304 stainless steel for engineering applications.

Keywords: AISI 304 stainless steel, Shot peening, Surface modification, Bending strength, Wear resistance, Corrosion-resistance and Microstructure.

1. INTRODUCTION

AISI-304 austenitic stainless steel is widely used owing to its excellent mechanical properties and corrosion resistance. Surface work hardening and grain refinement further improve fatigue strength and bending performance, making shot peening a widely used surface enhancement technique for fatigue-critical components in aerospace, automotive, railway, and power-generation industries [1–3]. Shot peening is a cost-effective surface treatment that improves hardness, fatigue, wear, and corrosion resistance through work hardening, grain refinement, and compressive residual stresses. The performance enhancement depends on shot peening parameters, particularly coverage, making parameter optimization essential. Optimum shot peening coverage is essential to maximize bending strength and fatigue life while avoiding excessive surface roughness and over-peening [5–7]. These findings indicate that corrosion behaviour depends on achieving an appropriate balance between microstructural refinement and surface scenery [8-9]. Although previous studies have reported improvements in individual properties, inadequate investigate has widely evaluated the combined possessions of shot peening exposure on bending strength, wear, corrosion behaviour, and microstructural evolution under identical conditions. This study investigates the influence of 100%, 150%, and 200% shot peening coverage to identify the optimum condition for maximizing the overall performance of AISI-304 stainless steel.

2. Properties of Tested Specimen

AISI-304 stainless steel is a versatile austenitic alloy acknowledged for its excellent corrosion resistance, high-quality mechanical properties, and formability. The chemical composition, mechanical properties, and proportions of the test specimens are obtainable in Table 2.1 and Figure 2.1 respectively.

Table 2.1 Properties of the tested specimen

Compo-nent	Chromi-um (%)	Carbon (%)	Iron (%)	Mangan-ese (%)	Nickel (%)	Silicon (%)	Phos-phorus (%)	Sulfur (%)	Density (g/cm³)	Tensile Strength (MPa)	Yield Strength (MPa)	Elongation at Break (%)	Modulus of Elasticity (GPa)
Stainless Steel (AISI-304)	18 – 20	Max 0.08	66.345-74	Max 2	8-10.5	Max 1	Max 0.045	Max 0.03	7.8–8.03	500–750	200–250	30–45	193

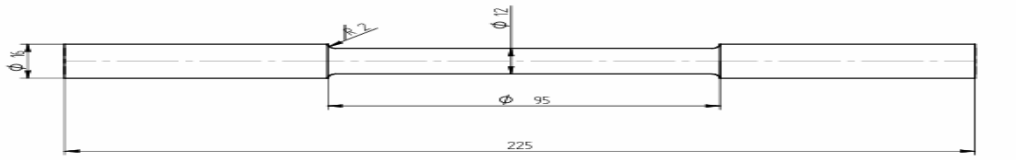


Figure 2.1 The dimensions of round bar specimen (AISI 304 Stainless Steel)

3. Experimental Methodology

AISI-304 stainless steel testing specimens were prepared according to standard dimensions and subjected to shot peening at 100%, 150%, and 200% coverage using a blast velocity of 31.49 m/s. The shot peening process induced surface plastic deformation and compressive residual stresses, while the peening intensity (0.21–0.45 mmA) was controlled using Type A Almen strips. The mechanical performance was evaluated throughout three-point bending tests using a computerized axial fatigue testing machine. Wear performance was assessed by means of a pin-on-disc wear tester by measuring weight loss under different normal loads. Corrosion resistance was investigated by immersing the specimens in a 20 wt. % NaCl solution for 15 days, with load measurements in use each 24 h to establish the percentage weight loss. The experimental procedure enabled a comparative evaluation of the persuade of shot peening exposure on the bending strength, wear resistance, and corrosion behaviour of AISI-304 stainless steel.



Figure 3.1 Fatigue Testing Machine

Table 3.1 Specification of Computerized Fatigue Testing Machine

Parameter	Specification
Capacity	50 KN
Type	Dynamic Servo Hydraulic Model
Max. Load	+ 50 KN
Max. Load for Dynamic Fatigue	+ 10 KN
Max. Displacement	+50 to 100 mm
Max.Frequency	2 HZ
Load resolution	1 KGF
Displacement Resolution	0.01 mm
Width Between column	200 mm
Max length of specimen	300 mm
Control system	PID-Based Controller
Hydraulic system	150 bar, 2 LPM system
Duct Cycle	40%
Oil Capacity	50Litres
Load cell type	Pan Cake
Displacement Sensor Type	High-resolution non-contact linear magnetic scale

4. Results and Discussion

Experimental investigations, including bending, wear, and corrosion tests, were performed on AISI-304 stainless steel among non-shot peened and shot peened (100%, 150 % and 200 % coverage), respectively .

4.1 Analysis of bending Strength with and without shot peened process

Figure 4.1 presents the load–deflection curves of untreated (NSP) and shot-peened AISI-304 stainless steel specimens. Shot peening appreciably enhanced the bending performance compared with the untreated specimen. The 100% shot-peened specimen exhibited a moderate increase in load-carrying capacity, while the 150% and 200% shot-peened specimens showed the highest bending strength due to surface work hardening and compressive residual stresses. Among the tested conditions, 150% shot peening coverage demonstrated the maximum load-bearing capacity, with an improvement of approximately 18–20% over the untreated specimen. Overall, the outcome substantiate that shot peening efficiently enhances the flexural behaviour of AISI-304 stainless steel by delaying crack initiation and improving resistance to plastic deformation.

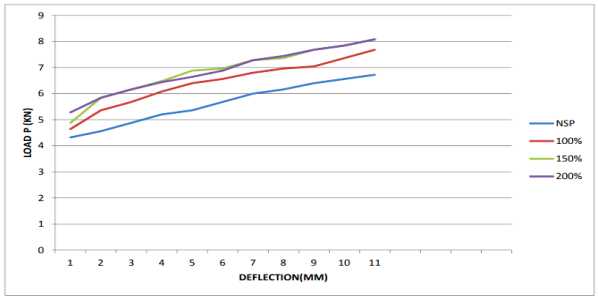


Figure 4.1 Load–Deflection Curves of Untreated and Shot-Peened AISI-304 Stainless Steel Specimens

4.2 Comparison of Wear Behaviour between Untreated (NSP) and 150% Shot-Peened AISI-304 Stainless Steel

Figure 4.2 compares the weight loss and wear rate of untreated (NSP) and 150% shot-peened AISI -04 stainless steel under normal loads of 1000 g, 2000 g, and 3000 g. The untreated specimen exhibited privileged weight loss with rising load, whereas the 150% shot-peened specimen showed significantly lower material loss due to surface hardening, grain refinement, and compressive residual stresses induced by shot peening. The wear rate amplified with load for both specimens, but the shot-peened specimen maintained a more stable wear response at higher loads. Overall, the results demonstrate that 150% shot peening effectively enhances the wear resistance of AISI-304 stainless steel by reducing material removal during sliding wear.

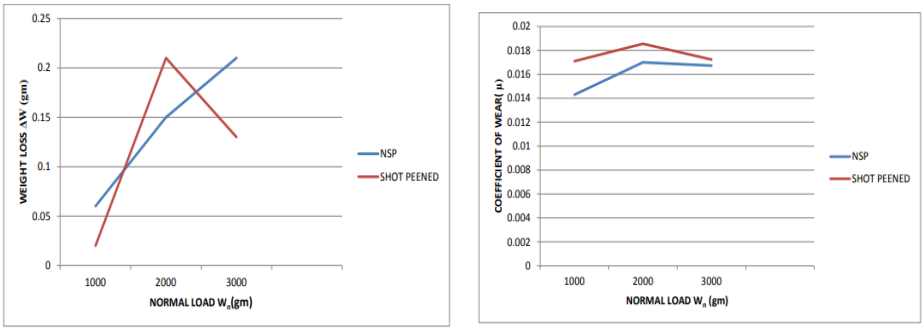


Figure 4.2 Compares the weight loss and **wear rate** of untreated (NSP) and **150% shot-peened** AISI-304 stainless steel specimen

4.3 Variation of Specimen weight with exposure time for untreated and Shot-Peened specimen

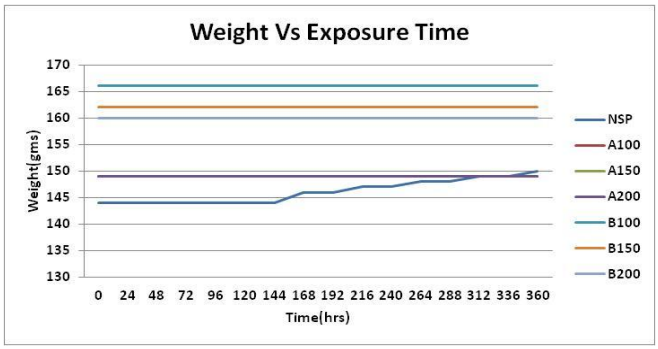


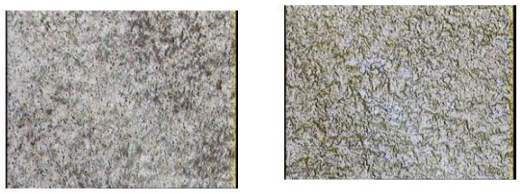
Figure 4.3 Variation of Specimen Weight with Exposure Time for NSP and Shot-Peened specimen

Figure 4.3 shows the corrosion behaviour of untreated (NSP) and shot-peened AISI-304 stainless steel during a 360-hour salt exposure test. The untreated specimen exhibited a gradual weight increase after 168 hours owing to the configuration of corrosion products, reaching approximately 6 g of weight gain after 360 hours. In contrast, all shot-peened specimens (100%, 150%, and 200%) maintained nearly constant weights throughout the test, indicating negligible corrosion. The improved corrosion resistance is attributed to the formation of compressive residual stresses, grain refinement, and a stable passive film induced by shot peening. The results

demonstrate that shot peening considerably enhances the corrosion resistance of AISI-304 stainless steel compared with the untreated condition.

4.4 Microstructure Analysis of Non-Shot-Peened (NSP) and 150% Shot-Peened AISI-304 Stainless Steel

Figure 4.4 compares the optical microstructures of untreated (NSP) and 150% shot-peened AISI-304 stainless steel at 250× magnification. The untreated specimen exhibited a coarse, equiaxed austenitic grain structure, whereas the 150% shot-peened specimen showed significant grain refinement, deformation bands, and strain-induced martensite due to severe plastic deformation. These microstructural changes enhanced surface work hardening, resulting in improved hardness and superior mechanical, wear, and corrosion performance.



a) Austenite Microstructure b) Martensite Microstructure

Figure 4.4 Optical Microstructures of Non-Shot-Peened (NSP) and 150% Shot-Peened AISI-304 Stainless Steel

5. Conclusions

The present study demonstrates that 150% shot peening significantly enhances the performance of AISI 304 stainless steel by improving bending strength, wear resistance, and corrosion resistance. These improvements are attributed to surface work hardening, grain refinement, deformation bands, strain-induced martensite, and beneficial compressive residual stresses. The refined surface microstructure effectively reduces material removal and promotes the formation of a stable passive layer. Overall, 150% shot peening is an effectual surface treatment for increasing the durability, reliability, and service life of AISI-304 stainless steel components operating under cyclic loading, wear, and corrosive environments.

Acknowledgement

We sincerely acknowledge the support of VGST (Karnataka Science and Technology Promotion Society – KSTePS), Government of Karnataka, in collaboration with DST, Government of India, for their continued assistance and for providing access to research equipment.

References

1. Nalla, R. K., Altenberger, I., Noster, U., Liu, G. Y., Scholtes, B., & Ritchie, R. O. (2003). On the influence of mechanical surface treatments—deep rolling and laser shock peening—on the fatigue behavior of Ti–6Al–4V at ambient and elevated temperatures. Materials Science and Engineering: A, 355(1-2), 216-230.
2. Keshavamurthy, R., Muneer, S. A. R., Nirmala Devi, E., Naga Sai, A., Ram Prasad, C., Satyanarayana, M. V. N. V., & Venkata Ramana, V. S. N. (2025). Influence of shot peening on microstructure and electrochemical behavior of al-Cu alloy. Canadian Metallurgical Quarterly, 1-12.
3. Haribaskar, R., & Sampath Kumar, T. (2024). Surface integrity of consecutive shot-peened additively manufactured 316L stainless steel. Materials and Manufacturing Processes, 39(13), 1817-1829.
4. Qin, Z., Li, B., Zhang, H., Wilfried, T. Y. A., Gao, T., & Xue, H. (2022). Effects of shot peening with different coverage on surface integrity and fatigue crack growth properties of 7B50-T7751 aluminum alloy. Engineering Failure Analysis, 133, 106010.
5. Závodská, D., Guagliano, M., Bokůvka, O., & Trško, L. (2016). Fatigue resistance of low alloy steel after shot peening. *Materials Today: Proceedings*, *3*(4), 1220-1225.
6. Świetlicki, A., Szala, M., & Walczak, M. (2022). Effects of shot peening and cavitation peening on properties of surface layer of metallic materials—a short review. *Materials*, *15*(7), 2476.
7. Chung, Y. H., Chen, T. C., Lee, H. B., & Tsay, L. W. (2021). Effect of micro-shot peening on the fatigue performance of AISI 304 stainless steel. *Metals*, *11*(9), 1408.
8. Lee, H. S., Kim, D. S., Jung, J. S., Pyoun, Y. S., & Shin, K. (2009). Influence of peening on the corrosion properties of AISI 304 stainless steel. *Corrosion science*, *51*(12), 2826-2830.
9. Bagherifard, S., Slawik, S., Fernández-Pariente, I., Pauly, C., Mücklich, F., & Guagliano, M. (2016). Nanoscale surface modification of AISI 316L stainless steel by severe shot peening. *Materials & Design*, *102*, 68-77.

A Machine Learning Approaches for Left Ventricular Hypertrophy Detection

Chaitanya V[1], Lavanya N L[2], Nirmala S Gupta[3]

[1]*BMS Institute of Technology, chaitanya.v@bmsit.in*
[2]*East West College of Engineering, lavanyaanand96@gmail.com*
[3]*Sri Krishna Institute of Technology, nirmala.sguptha@gmail.com*

Abstract

Left Ventricular Hypertrophy is a cardiac anatomical problem characterized by left ventricular hypertrophy that results from hypertension, valve disease, and increased risk in cardiology. Early detection is critical for proper treatment and avoiding any future complications. Traditional diagnostic methods such as echocardiography and cardiovascular magnetic resonance imaging can be described as old standard methods that are quite expensive and impractical in population screening. Diagnostic criteria based on electrocardiogram analysis can also prove to be cost-effective and easy to access; however, problems related to low sensitivity and specificity could occur. This paper seeks to review and analyse algorithmic analysis and deep learning approaches for LVH detection and evaluate their possible advantages over traditional methods. Recent literature on machine learning and deep learning approaches for LVH classification using large-scale ECG data sets, imaging data, and multimodal clinical data. However, the emphasis shall be laid on the advancements in features extraction, classification and explainability of models-classical machine learning techniques, CNNs and transformers. Advancements in machine learning/deep learning algorithms have demonstrated great potential in improving the efficacy of LVH detection methods. They have yielded improved sensitivity and specificity as opposed to traditional ECG-based methods despite the issues related to class imbalance, heterogeneous nature of data. The combination of machine learning techniques with the field of clinical cardiology would be highly useful in making the design of LVH detection models more aesthetically appealing.

Key words. Left Ventricular Hypertrophy, Machine Learning, Deep Learning, ECG, Medical Imaging, Cardiology

1. INTRODUCTION

Left Ventricular Hypertrophy is a medical condition that exists within the myocardium that involves the thickening of the muscle fibres in the left ventricle owing to the occurrence of prolonged high blood pressure, problems with valves, or other kinds of cardiomyopathies. Numerous scientific studies have shown that people who suffer from LVH have a much greater risk of developing diseases such as heart failure, arrhythmia, stroke, and sudden cardiac death. Conventional methods of diagnosing LVH through imaging alone, such as echocardiography and cardiac magnetic resonance imaging, remain the gold standard. but, these modalities are resource-intensive, requiring specialized equipment, skilled employees, and sizeable time, thereby proscribing their feasibility for massive populace-stage screening. however, the electrocardiogram is inexpensive, non-invasive, and extensively to be had in clinical exercise, but traditional ECG-primarily based diagnostic standards consisting of the Sokolow-Lyon index and Cornell voltage criteria suffer from low sensitivity and restricted generalizability across various affected person populations [5]. These limitations have fuelled a developing interest in the use of machine getting to know ML and deep learning methods to improve LVH detection via leveraging each imaging and ECG statistics at scale.

In recent years, large developments have been made via the utility of synthetic intelligence to cardiovascular diagnostics. EchoNet-LVH represents a primary advance in video-based echocardiographic analysis by using convolutional neural networks on over 28,000 echocardiogram motion pictures to are expecting wall thickness and underlying etiologies, reaching near-professional performance with an AUC 0.98 for hypertrophic cardiomyopathy detection[5].Even as this confirmed the feasibility of automatic video interpretation, its reliance on echocardiography despite the fact that poses demanding situations for large-scale deployment. In parallel, Kokubo et al [2] evolved a deep gaining knowledge of model skilled on 12-lead ECG waveforms, reporting an AUROC of 0.784 for LVH detection in a large Japanese dataset, consequently highlighting the functionality of ECG-primarily based AI techniques in scientific screening. Complementary research, such as Zhao et al [4].'s CNN-LSTM hybrid [21] further showcased the cost of mixing spatial and temporal information from ECG indicators, on the same time as blended ECG preprocessing with function engineering and ML classifiers to decorate ordinary overall performance in assessment to conventional voltage-based completely criteria. Feature-based approaches, like those [3] applied in C5.0 decision tree, and by Tacconne et al.[13] in Random Tree and SVM methods in PTB-XL and Georgia ECG datasets, have shown to be promising internally but have faced challenging situations in terms of external generalization, highlighting the need for cross-cohort validation. Further advancements in the field have involved interest-based approaches, that include, Ryu et al[6] CoAt-Mixer model, which involved convolution and self-attention to identify dispersed sex-specific markers in ECG associated with LVH.

2. LIMITATIONS OF EXISTING SURVEYS

- Many surveys check each ECG-based or imaging-based LVH detection techniques in isolation, developing a fragmented view of the sector.
- Several opinions rely on limited or single-organisation datasets, limiting their capability to evaluate model

nor-mal overall performance throughout several populations.

- Comparisons among classical ML and DL strategies are often high level, with insufficient evaluation of architectures, strategies, or evaluation constraints.
- Problems including dataset imbalance, demographic variability, and inconsistent diagnostic labelling are not often noted intensive, despite their primary effect on model reliability.
- Few studies compare the scientific interpretability of ML models or deal with the gap among set of regulations performance and actual-global deployment.
- Emerging tactics—which include transformer-based to-tally models, attention mechanisms, and multimodal analysing are underrepresented in many gift surveys.
- Most critiques do no longer provide unified benchmarking or standardized metrics, making it difficult to have a look at effects throughout great studies.

A. Purpose of the Survey

Left Ventricular Hypertrophy is an important cardiac disorder that is related to increased morbidity and mortality risk. The conventional diagnostic techniques that include echocardiography and cardiac MRI remain the gold standard for LVH diagnosis but due to their high cost and limited availability, they cannot be used for screening on a large scale. With the fast development of system analysis and deep learning algorithms [20], more and more research papers are going to appear discussing the potential role of the above-mentioned approaches in enhancing the accuracy and effectiveness of LVH diagnosis. The goal of the present literature review is to analyze current machine-learning techniques in the field of cardiology.

B. Challenges and Limitations

Despite all the potential benefits, many difficult conditions exist that can prevent the implementation of ML based LVH diagnosis. These may be described as:

- Unbalanced classes: In any scientific dataset, there will always be more instances of normality than pathology influencing the bias to favor the accuracy of the model [2][3]
- Data variability: The difference in recording equipment, demographic groups, and scientific methods makes it difficult for the model to be generalized [11]
- Black box models: The lack of an explanation on how decision making is done by most deep learning algorithms reflects negatively on clinicians' confidence [14].
- Resource requirement: A great need for computation and large annotated data sets hinders implementation in low-resource areas [20].
- Limited clinical validation: Few models have been tested in clinical settings to determine their reliability in clinical settings.

C. Contribution

The present review offers the following contributions:

1. Provides an analysis of the evaluation of ML and DL approaches applied to LVH detection on modalities that include ECG and imaging data.
2. Provides comparative analysis of classical ML models, convolutional neural networks, and transformer-based approaches regarding their advantages, disadvantages, and scientific application.
3. Emphasizes key challenges that consist of class imbalance, data heterogeneity, and model interpretability and gives insights into potential ways to solve them; points out gaps in existing literature and directions for future research to create robust and clinically reliable LVH detection models.
4. Points out gaps in existing literature and directions for future research in order to create robust and clinically reliable LVH detection models.

3. SURVEY METHODOLOGY

To create a targeted and reproducible survey of device analyzing and deep getting to know techniques for Left Ventricular Hypertrophy detection, we determined a scientific method that synthesizes proof from representative research spanning ECG-, echo-, and multimodal-based techniques. The evaluation approach involved (a) choice of landmark and brand-new papers (collectively with EchoNet-LVH and vital ECG/ML contributions), (b) extraction of look at tendencies (modality, dataset duration, model structure, enter type, pronounced metrics, and boundaries), and (c) thematic synthesis to answer focused studies questions and map scientific manifestations to modeling picks.

A. Symptoms for Causing Left Ventricular Hypertrophy

The following clinical conditions and symptoms are commonly associated with the development of LVH:

1. Chronic Hypertension is the main reason for LVH because the high blood pressure makes the left ventricle work harder against the raised afterload, making it become thick due to the compensation process [1][2].
2. Valvular Heart Disease, especially diseases such as aortic stenosis or aortic insufficiency, raises the pressure and volume load on the left ventricle, causing LVH. Hypertrophic Cardiomyopathy in addition contribute to LVH through strange and disproportionate thickening of the myocardium independent of hemodynamic strain.

3. Weight problems and Metabolic Syndrome raise cardiac workload and metabolic call for, predisposing individuals to structural cardiac variations [13].
4. Athlete's coronary heart represents a physiological and commonly benign form of LVH that arises from lengthy-time period in depth physical training [15].
5. Shortness of Breath (Dyspnea) usually arises for the duration of exertion because of impaired ventricular compliance and diastolic sickness [13].
6. Palpitations and Arrhythmias, including atrial fibrillation, which frequently accompany structural remodeling.
7. Syncope or Dizziness as a result of poor cardiac output and arrhythmic attacks
8. Tiredness and poor exercise tolerance, which can be results of poor ventricular filling and inefficiency of cardiac output production.

4. RELATED WORK

This section summarizes preceding research on LVH detection the use of device studying and deep learning methods, highlighting methodologies, creator contributions, dataset utilization, and guide tendencies.

Research Methodology: This section summarizes previous studies on LVH detection using tool studying and deep learning techniques, highlighting methodologies, author contributions, dataset utilization, and guide inclinations, data preprocessing, characteristic extraction, and validation techniques range, with a few research the usage of internal cross-validation and others reporting outside check set performance. These studies employed a retrospective cohort look at layout to develop and validate a device getting to know version for Left Ventricular Hypertrophy (LVH) detection. The observe become performed in 3 stages: records collection and preprocessing, model development, and validation.

TABLE I
COMPARISON OF LVH DETECTION METHODS (EXAMPLE METRICS)

Method	AUROC	Sensitivity	Specificity
Cornell + Sokolow–Lyon	0.65	0.30	0.90
SVM (engineered features)	0.75	0.60	0.80
CNN on raw ECG	0.85	0.72	0.82
Transformer + demographics	0.90	0.78	0.85

TABLE II
PERFORMANCE CONTRAST OF ML ALGORITHMS FOR LVH DETECTION

Model	Accuracy	Precision	Recall	Specificity	F1-Score	ROC-AUC
Logistic Regression	85.2%	83.4%	81.7%	87.1%	82.5%	0.89
Random Forest	91.0%	90.2%	88.9%	92.4%	89.5%	0.94
SVM (RBF Kernel)	88.7%	87.1%	86.3%	89.8%	86.7%	0.92
CNN (ECG signals)	94.3%	93.5%	92.8%	95.0%	93.1%	0.97

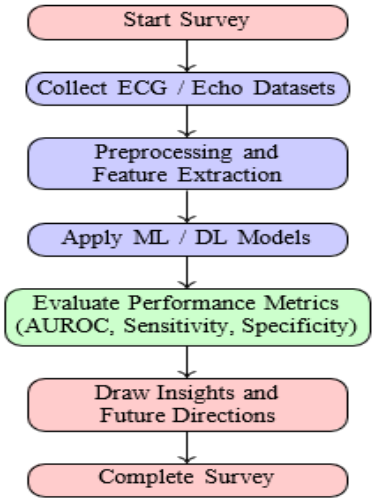


Fig. 2. Step-by-Step Pipeline of the Survey Approach.

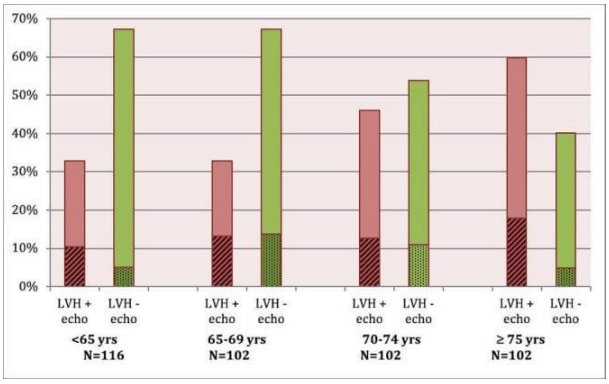


Fig. 3. Summarizes the performance of different algorithms on the test dataset

5. FUTURE WORK

The future of LVH detection involves improving the LVH detection models by employing multi-centre datasets, which involve a variety of patients. This will make the detection models more robust and generalize better. The use of explainable AI models and clinician-in-the-loop systems will help in making the models more transparent. Calibration methods like temperature scaling and conformal prediction will facilitate safe decision-making in the clinic. Class imbalance will be addressed using data augmentation, generative adversarial networks, and cost-sensitive learning.

6. CONCLUSION

The survey covered current developments in LVH detection through device-based analysis and application of ML methods. The evolution of conventional models that combine features with logistic regression, decision trees, and SVM techniques has been directed towards more sophisticated model architectures including CNNs, RNNs, and hybrid models. AI-based systems substantially increase sensitivity, specificity, and diagnostic

accuracy in comparison with conventional ECG and echocardiography criteria. With the use of large data sets (PTB-XL, UK Biobank, and ECG/echo data banks from various organizations), CNN techniques produce AUROC scores of more than 0.95 for the positive samples. Attention and hybrid models provide an additional advantage due to the use of temporal and morphological characteristics. Nevertheless, some challenges exist such as limited population data, no external validation, and bias in the data sets. To deploy the models in clinical practice, there is need to develop models that can generalize across populations, imaging modalities, and other co-morbid conditions. Explainability, transparency, and regulatory approval are among the challenges in deploying the model in clinical practice. The answers to such challenges could come through system-level studies and designing multi-modal systems incorporating ECG and echocardiographic and clinical data.

REFERENCES

[1] Duffy E.I., et al. "EchoNet-LVH: High-Throughput Preci-sion Phenotyping of LVH with Cardiovascular Deep Learn-ing," *Nature Medicine / arXiv*, 2021–2022. Available at: https://arxiv.org/abs/2111.10425

[2] Kokubo A., et al. "Automatic Detection of Left Ventricular Dilatation and LVH from 12-lead ECG," *International Heart Journal*, 2022. Available at: https://www.jstage.jst.go.jp/article/ihj/63/1/63₂1 − 275/*article*

[3] De la Garza Salazar J., et al. "Optimizing ECG to Detect Echocardiographic LVH with ML," *PLOS ONE*, 2021. Available at: https://journals.plos.org/plosone/

[4] Zhao L.T., et al. "CNN-LSTM Model for LVH Detection from 12-lead ECG," *Frontiers in Cardiovascular Medicine*, 2022. Available at: https://www.frontiersin.org/journals/cardiovascular-medicine

[5] Liu C.W., et al. "ECG Signal Processing and Machine Learn-ing for LVH Detection," *Scientific Reports*, 2023. Available at: https://www.nature.com/srep/

[6] Ryu J.S., et al. "Self-Attention Deep Learning (CoAt-Mixer) for LVH Detection from ECG," *PLOS ONE*, 2023. Available at: https://journals.plos.org/plosone/

[7] Naderi H., et al. "ECG Biomarkers from UK Biobank to Predict Echo-LVH," *EHJ Digital Health*, 2023. Available at: https://academic.oup.com/ehjdh

[8] Liu C.M., et al. "AI-enabled Model for Early Detection and Risk Stratification of LVH," *Circulation: Outcomes*, 2022. Available at: https://www.ahajournals.org/journal/circoutcomes

[9] Rabkin S.W., et al. "Searching for the Best ML Algorithm for LVH Detection: A Systematic Review," *MDPI Biosensors*, 2024.
Available at: https://www.mdpi.com/journal/biosensors

[10] Seranton S. "Systematic Review and Meta-analysis of AI for LVH Detection," *Scientific Reports*, 2024. Available at: https://www.nature.com/srep/

[11] Hybrid LSTM Model. "Hybrid LSTM Deep Learning for LVH Classification," Institutional report, 2024. Available at: https://scholar.google.com/

[12] Dwivedi T., et al. "ML Models on 6-lead ECGs for LVH In-terpretation," *Journal of Electrocardiology*, 2023. Available at: https://www.jecgonline.com/

[13] Taccone J., et al. "ECG-based LVH Detection using PTB-XL and Georgia Datasets with RF/SVM," Preprint, 2024. Available at: https://arxiv.org/

[14] MedRxiv Preprint. "Benchmarking ML Models (LGBM, ResNet, MLP) for LVH Detection," 2024. Available at: https://www.medrxiv.org/

[15] Research Preprint. "An ECG-based Model for LVH Detection (Feature-based RF/SVM)," Preprints, 2023–2024. Available at: https://www.preprints.org/

[16] Gupta A., et al. "Comparison of machine learning models for LVH detection using VAE-based ECG feature extraction," *Scientific Reports*, 2024. Available at: https://www.nature.com/srep/

[17] Huang Y., et al. "Artificial intelligence-based LVH detection using CatBoost on ECG data," *European Heart Journal – Digital Health*, 2025. Available at: https://academic.oup.com/ehjdh

[18] Srinart N., et al. "Artificial intelligence for the electrocardiographic diagnosis of LVH: Systematic review and meta-analysis," *Scientific Reports*, 2024. Available at: https://www.nature.com/srep/

[19] Liu C.M., et al. "Artificial Intelligence-enabled Electrocar-diography for Detection and Risk Stratification of LVH," *Circulation: Cardiovascular Outcomes*, 2022. Available at: https://www.ahajournals.org/journal/circoutcomes

[20] M. S, J. S. Jakati, C. Umarani. "Deep Learning Methods for Liver Tumor Segmentation: An Extensive Investigation and Analysis" *1st International Conference on Advancement in Futuristic Technologies (ICAFT)*, 2025. Available at: doi: 10.1109/ICAFT66710.2025.11452940

[21] Pradeep Kumar, B.P, et al. "A Comparison of CNN, RNN, and FNN Algorithms to Investigate Effective Diabetes Pre-diction" *Arch Computat Methods Eng* , 2025. Available at: https://doi.org/10.1007/s11831-025-10467-6

IC-DAIS 2026
<Conference Series name> <Volume number and Year> <DOI Number>

AI-Based Embodied Carbon Estimation from Bill of Quantities Data in Indian Construction Projects

Ankith Prabhu

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India, eng23am0286@dsu.edu.in

Ritika Reddy Alla

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India, eng23am0289@dsu.edu.in

Shreya Rao

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India, eng23am0278@dsu.edu.in

Ankita Thakkar

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India, ankitath1989@gmail.com

Yisol Choi

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India, eng23am0285@gmail.com

Lakshmanan M

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India,, lakshmanan1909@gmail.com

Abstract.

No practical method exists for estimating embodied carbon at the Bill of Quantities (BOQ) stage even though construction projects produce significant amounts of it. This paper proposes using BOQ data from five real Indian construction projects to directly predict Carbon Cost Intensity (CCI) which is the cost per unit of embodied carbon. Emission factors are drawn from ICE DB V4.1. Item descriptions are vectorized using TF-IDF, combined with one-hot encoded material and unit categories, concrete grade, and log-transformed quantity. Five regression models were trained and evaluated under 5-fold cross-validation: XGBoost, Support Vector Regression, Decision Tree, Random Forest, and Linear Regression. XGBoost performed best, achieving a mean absolute error of 37.56 INR/kgCO2e and an R^2 of 0.594. A GGBS substitution analysis showed that replacing 30% and 50% of cement with GGBS cut total portfolio carbon by 24.1% and 40.2% respectively. The outcome is a working, reproducible framework for embodied carbon assessment built around the realities of Indian construction practice.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Keywords. embodied carbon, bill of quantities, machine learning, XGBoost, carbon cost intensity, GGBS, Indian construction, emission factors, ICE DB

1. INTRODUCTION

Buildings and construction account for roughly 37% of global energy and process-related CO_2 emissions [1], while embodied carbon is largely determined during design and procurement [2]. In India, standardized project-level assessment tools remain limited, and conventional LCA requires detailed material inventories often unavailable at the BOQ stage [3,4]. We therefore investigate whether BOQ data can directly support early-stage embodied carbon estimation using Carbon Cost Intensity (CCI), defined as cost per unit of embodied carbon. This study develops a BOQ-based ML pipeline using ICE DB V4.1 emission factors, evaluates regression models for CCI prediction, and examines portfolio-level carbon reductions under GGBS substitution scenarios.

2. RELATED WORK

Embodied carbon estimation falls under process-based LCA, input–output analysis, and hybrid methods, with machine learning gaining attention as a data-driven route [7], as in Pomponi et al. [8] and Kayaoglu et al. [9]; neither confronts the inconsistent unit conventions typical of Indian BOQs. Factors here are sourced from ICE DB V4.1 [10], and XGBoost [11] is generally the strongest choice for tabular data of this kind.

3. DATASET AND METHODOLOGY

The dataset comprised 1161 BOQ items from five Indian construction projects. Descriptions and units were standardized, while material categories and concrete grades were extracted using keyword-based rules. PA (321 rows) is the only priced BOQ; CCI = Unit Rate (INR) / Emission Factor (kgCO2e/unit), using ICE DB V4.1 values (e.g., M20 concrete: 284.9, steel Fe500: 1.720 kgCO2e/kg). Raw CCI is heavily right-skewed (skewness = 6.95); log transformation reduces this to 0.49, and models predict log(CCI), converted back before evaluation. Features combined TF-IDF representations of item descriptions with unit, material category, concrete grade, and log-transformed quantity.

Five models were compared — XGBoost, SVR, Decision Tree, Random Forest, and Linear Regression — alongside Ridge, Lasso, and an MLP. All went through 5-fold CV stratified by CCI quartile, with MAE/RMSE/R^2 reported on the original scale. A leave-one-project-out test also checked cross-project transfer, since all labeled data comes from the single priced source, PA.

4. RESULTS AND DISCUSSION

Table I shows cross-validated results for all seven models. XGBoost performed best (MAE 37.56 INR/kgCO2e, R^2 0.594); SVR came closest (MAE 39.04, R^2 0.546). Linear models and the MLP generalized worse, with the MLP yielding a negative R^2. Performance is most sensitive to items lacking a clear material keyword and to pricing outliers in PA, but the MAE falls within 12–18% of median CCI, within the tolerance expected of early-stage Indian estimates.

TABLE I
5-FOLD CROSS-VALIDATION RESULTS

Model	MAE (INR/kgCO2e)	RMSE (INR/kgCO2e)	R^2
XGBoost	**37.56**	**89.18**	**0.594**

SVR	39.04	94.30	0.546
Lasso Regression	53.35	109.89	0.384
Decision Tree	45.26	111.57	0.365
Ridge Regression	60.14	117.68	0.293
Random Forest	54.03	118.37	0.285
Neural Network (MLP)	64.72	171.02	−0.493

Material category and unit type account for over 60% of XGBoost's prediction gain. CCI was predicted for 577 line items across the four unpriced projects; concrete and steel together account for over 80% of predicted embodied carbon, while items with no ICE DB match are excluded from totals.

5. GGBS Substitution Analysis

GGBS (Ground Granulated Blast-furnace Slag) can reduce embodied carbon in Portland cement concretes [12]. Three portfolio-wide cases were studied: Baseline (no GGBS), 30% GGBS (workable without affecting structural performance [13]), and 50% GGBS (practical for less demanding elements). The adjusted emission factor is:

$$EF\_blend = EF\_baseline \times (1 - f\_GGBS) + EF\_GGBS \times f\_GGBS,\ \ EF\_GGBS = 0.083\ kgCO2e/kg$$

TABLE II
PORTFOLIO-LEVEL GGBS SCENARIO RESULTS

Scenario	GGBS (%)	Total Carbon (tCO2e)	Reduction (%)	Cost Premium (INR M)
A: Baseline	0	167,109	—	0.0
B: 30% GGBS	30	126,844	24.1	34.1
C: 50% GGBS	50	100,001	40.2	56.8

A 30% GGBS mix removes 40,265 tCO2e (24.1%) for INR 34.1 million; 50% saves 67,109 tCO2e (40.2%) at INR 56.8 million. Most savings are driven by M20 and M30 concrete. Carbon saved per INR million is nearly identical (~1181-1182 tCO2e) at either level. Even a conservative 30% shift delivers substantial decarbonization at a cost well within standard contingencies, with no redesign or new materials required.

6. Limitations and Future Work

All 194 labeled samples come from one priced project (PA), so it is unclear whether the learned CCI patterns generalize to other regional cost environments. So single sourcing is the most significant constraint. A leave one project out test offers only a partial check. ICE DB coverage is also incomplete, so the totals are lower bounds.

7. Conclusion

This pipeline estimates embodied carbon from BOQ data in Indian construction projects, mapping ICE DB V4.1 emission factors to line items and combining TF-IDF text features with structured attributes to estimate CCI without a full life cycle assessment. XGBoost performed best among seven models (MAE 37.56 INR/kgCO2e, R^2 0.594). The GGBS result is the most useful takeaway: replacing just 30% of cement with GGBS cuts portfolio

carbon by 24.1% with no redesign and a cost within typical Indian project contingencies. How well the model travels to other sites remains open — a question more priced BOQs should help resolve.

8. REFERENCES

[1] United Nations Environment Programme, "2022 Global Status Report for Buildings and Construction," UNEP, Nairobi, 2022.

[2] F. Pomponi and A. Moncaster, "Embodied carbon mitigation and reduction in the built environment: What does the evidence say?," Journal of Environmental Management, vol. 181, pp. 687–700, Oct. 2016.

[3] K. G. Akshatha, S. Palaniappan, and M. K. Pradeep Kumar Rao, "Embodied carbon variability in Indian high-rise residential buildings: A case-based assessment," in Proc. 13th World Construction Symposium, Sri Lanka, Aug. 2025, pp. 623–634.

[4] C. De Wolf, F. Pomponi, and A. Moncaster, "Measuring embodied carbon dioxide equivalent of buildings: A review and critique of current industry practice," Energy and Buildings, vol. 140, pp. 68–80, Apr. 2017.

[5] Z. H. Ali and A. M. Burhan, "Hybrid machine learning approach for construction cost estimation: an evaluation of extreme gradient boosting model," Asian Journal of Civil Engineering, vol. 24, pp. 2427–2442, 2023.

[6] X. Zhang et al., "Predictive models of embodied carbon emissions in building design phases: Machine learning approaches based on residential buildings in China," Building and Environment, vol. 258, 2024.

[7] A. Chadha and A. Gupta, "Role of machine learning to analyze the impact of construction industry on the carbon footprint," SSRG International Journal of Civil Engineering, vol. 12, no. 8, pp. 176–183, 2025.

[8] F. Pomponi, C. De Wolf, and A. Moncaster, Eds., Embodied Carbon in Buildings: Measurement, Management, and Mitigation. Cham: Springer, 2018.

[9] M. Kayaoglu, O. Tatari, and C. Skibniewski, "Machine learning for construction cost and embodied carbon estimation," Journal of Construction Engineering and Management, vol. 148, no. 3, 2022.

[10] C. Hammond and C. Jones, Inventory of Carbon and Energy (ICE) Database V4.1. Bath: University of Bath, 2019.

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, 2016, pp. 785–794.

[12] K. L. Scrivener, V. M. John, and E. M. Gartner, "Eco-efficient cements: Potential economically viable solutions for a low-CO2 cement-based materials industry," Cement and Concrete Research, vol. 114, pp. 2–26, Dec. 2018.

[13] Bureau of Indian Standards, IS 456:2000 Plain and Reinforced Concrete – Code of Practice, 4th rev. New Delhi: BIS, 2000.

<IC-DAIS>
<Conference Series name> <Volume number and Year> <DOI Number>

Prediction of Cancer: By Fuzzy AHP and TOPSIS Methods

Trupthi Rao

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India trupthirao-aiml@dsu.edu.in

Ankita Thakkar

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India ankita.t-aiml@dsu.edu.in

Navya Holla K

Dept. of ECE, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India hollanavya@gmail.com

Dr. Manasa

Dept. of ECE, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India manasar.87@gmail.com

Shridhar B Devamane

Dept. of CSE, MS Engineering College, Bangalore, Karnataka, India devamaneshridhar@gmail.com

Gargi N

Dept. of AI &DS, Global Academy of Technology, Bangalore, Karnataka, India gargi@gat.ac.in

Abstract.

Cancer is considered as one of the most dangerous diseases around the world, it is the one were the abnormal cells rapidly starts growing causing the neighbouring cells or the neighbouring organs, where the abnormal cells are also called as the tumours. In this paper aims to predict the probability of cancer using some of the fuzzy methods like the Fuzzy AHP (Analytical Hierarchy Process) and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) methods. These methods are applied to a dataset which is taken from the Kaggle website, in this dataset includes attributes like smoking, genetic risk, chronic lung disease, air pollution, alcohol use or consumption. In this the first method used is the Fuzzy AHP method which ranks the importance of different criteria, here the smoking is the most important risk factor. The second method used is the TOPSIS method in this method the most important risk factor is smoking. This study helps us to understand that both the methods are effective in finding the cancer risk, while TOPSIS method is more advantageous over the AHP method for the quantitative analysis. The results show that smoking plays a most critical role in cancer prediction.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Keywords. Cancer prediction, cancer mortality, Fuzzy AHP, genetic predisposition, healthcare access, lifestyle factors.

1. INTRODUCTION

Cancer which is scientifically called as “Neoplasm” which basically means “New Growth” is one of the deadliest and life taking diseases in the world. The study and treatment of the cancer is generally called “Oncology” and the people responsible for treating this cancer are termed as “Oncologists”. [1] Cancer is mainly caused by carcinogenic cells also termed as the “Cancer Cells”. Such cells multiply in an exponential rate mainly causing tumours around the body organs which further on decrease the performance if that affected organ and even decrease the life expectancy of the organism affected. Such tumours are mainly classified into 2 types mainly Benign tumours and Malignant tumours. Among the above-mentioned tumours one of the dangerous tumours are “Malignant tumours” which are unlike “Benign tumours”. The main reason for their hazardousness is that they are not stagnant tumours like benign tumours and are constantly spreading all over the body and affecting the organs. There are more than 200 types of cancer present in the world. This study aligns with United Nations Sustainable Development Goal (SDG) 3, specifically Target 3.4 which aims to reduce premature mortality from non-communicable diseases including cancer.

2. RELATED WORK

Cancer is ranked as the 2nd most deadly disease responsible in loss of lives of many millions of people around the globe annually. It is said to believe that the life expectancy of a cancer patient is just around 5 years. [2] Based on the report produced by “The WHO” around 20 million new cancer cases are diagnosed with around 9.7 million deaths occurring annually all over the world. Based on the reports generated by “The International Agency for Research of Cancer (IARC)” which is official part of “The WHO” stated that “Pancreatic Cancer” was one of the major types of cancer that killed people much faster in an unprecedented rate and one of the most common types is the breast cancer in females. [3] In India around 850,00 people die annually to this deadly disease. Organizations like ICS, TMC and CPAA in India are constantly working in order to spread awareness on cancer and are even helping people in successfully treating cancer. [4] Based on the data reviewed by the WHO organization there has been an exponential increase in the deaths and has been a major concern in early diagnosis and prediction of whether a particular individual is affected by cancer. [10] The mortality rate due to cancer is around 146.0 per 1 lakh males and females. Around only 14% of cancers of the individuals living in the USA have been predicted. Some countries are unable to predict the cancer mainly due to 2 reasons: • Unaware of cancer disease. • Do not have the technology to detect cancer in an individual. [5] Due to the recent development in the areas of communication and other handheld gadgets, now it has become much easier to predict cancer in an individual. [6] Some of the different mobile applications mainly helping in such predictions are depicted in the Table I.

TABLE I. CANCER RELATED MOBILE APPLICATIONS

S. No.	Application Name	Description
1	Skin Vision	An AI based application detecting cancer by identifying lesions in the skin

2	Mole Mapper	Makes use of photographs and detects the moles and depicts the changes in Melatoma
3	. B4BC App	Mainly deals with the Breast cancer and helps in diagnosis
4	My Cancer Coach	Acts as a social app that helps in monitoring different individuals' stages of cancer
5	Breast cancer helpline	Connects users with other individuals facing breast cancer.

3. METHODOLOGY

We assume numerically for the pairwise comparison matrix as follows: Smoking: is 4 times more important than genetic risk. 3 times more important than Chronic lung disease. 6 times greater than air pollution. 5 times greater than alcohol use. Genetic risk: 3 times important than Chronic lung disease. 4 times important than air pollution. 3 times important than alcohol use. Chronic lung disease: 3 times important than air pollution. 4 times important than alcohol use. Air pollution: 4 times more important than alcohol use. The fuzzy geometric means are: Smoking (2.61,3.25,3.84), Genetic risk (1.19,1.55,1.93), Chronic Lung Disease (0.82,1.06,1.38), Air pollution (0.53,0.56,0.61), Alcohol use (0.31,0,33,0.38).

4. RESULTS

In fuzzy AHP method, based on the weights obtained we rank the alternative. The alternative with the highest weight is given rank 1 and so on as shown in Table II.

TABLE II. AHP RESULTS

Defuzzified Weights	Ranking
0.477	1
0.228	2
0.159	3
0.085	4
0.051	5

The value 0.477 with assigned rank value of 1 mainly corresponds to the attribute Smoking. Now moving forward with results of the TOPSIS method we again rank the closeness coefficients which are obtained as follows: [0.460, 0.357, 0.527] = [2, 3, 1]. A3 having the closeness coefficient of 0.527 is the best alternative. Now taking into consideration only the A3 alternative we get to know some key factors and can rank the criteria as follows: [Smoking, Genetic risk, Chronic Lung cancer, Air Pollution, Alcohol Use] = [1, 5, 2, 4, 3]. By the TOPSIS method too we get that smoking is one of the main causes for cancer.

5. CONCLUSION

We can conclude by Fuzzy AHP methods and TOPSIS methods "Smoking" is one of the most important criteria or the most cancer-causing factor present all over the world. By theory we can say that both the methods are equally important which is mainly based on the problem chosen and the type of data in it. The fuzzy AHP method performs very well if we have a complex decision problem with larger number of criteria to be considered and to be

ranked since AHP method can be used to handle such a large number of criteria by breaking them down into comparisons. Whereas the TOPSIS method mainly when we have a straightforward decision problem and even having quantitative data in it. In rare cases we have to use a combination of both these methods to obtain the best alternative. So thereby on the above-mentioned qualities of TOPSIS and AHP methods and by the dataset review discussed in the methodology section we get to know that AHP method serves as a best method to rank the alternatives. Hence "Smoking" is one of the important criteria to be noted down for cancer prediction.

6. REFERENCES

[1] N. K. Mudaliar, S. Redekar, N. M. Kadam and J. Ramachandran," Machine Learning-based Prediction Model for Cancer Detection," 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023.

[2] M. L. V. A. Priya and M. V. Subbarao," Performance Analysis of Machine Learning Techniques for Multi-Organ Cancer Detection and Classification: A Comparative Study," 2023 International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 2023.

[3] J. Liu, L. Fan, Q. Jia, L. Wen and C. Shi," Early Diabetes Prediction Based on Stacking Ensemble Learning Model," 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 2021.

[4] S. Mahajan, P. K. Sarangi, A. K. Sahoo and M. Rohra," Diabetes Mellitus Prediction using Supervised Machine Learning Techniques," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2023

[5] O. H. Kesav and R. G. K," A Systematic Study on Enhanced Deep Learning Based Methodologies for Detection and Classification of Early-Stage Cancers," 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Hamburg, Germany, 2023.

[6] R. D. Karthikeyan, R. G, V. V, G. B. C and K. M," A Review of Lung Cancer Detection using Image Processing," 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2021.

[7] G. Neelima, P. Kanchanamala, A. Misra and R. A. Nugraha," Detection Of Breast Cancer Based on Fuzzy Logic," 2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS), Bali, Indonesia, 2023.

[8] E. Hossain and M. A. Rahaman," Bone Cancer Detection & Classification Using Fuzzy Clustering & Neuro Fuzzy Classifier," 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT), Dhaka, Bangladesh, 2018.

[9] F. Taher and R. Sammouda," Lung cancer detection by using artificial neural network and fuzzy clustering methods," 2011 IEEE GCC Conference and Exhibition (GCC), Dubai, United Arab Emirates, 2011.

[10] Darshan, N., Darshan, S. R., Kodipalli, A., & Rao, T. (2024, April). Comparison of Various Computational Models for the Accurate Detection of Lung Cancer and Interpreting Using XAI. In 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS) (Vol. 1, pp. 1-6). IEEE.

<IC-DAIS>
<Conference Series name> <Volume number and Year> <DOI Number>

Integrated Fuzzy TOPSIS and AHP Approach for Selecting Car Brand using Multiple Decision Features

Trupthi Rao

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India trupthirao-aiml@dsu.edu.in

Souramita Bhowmik

Dept. of CSE (DS), Dayananda Sagar University, Bangalore, Karnataka, India souramita-ds@dsu.edu.in

Shivnandan Rai

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India shivnandanrai-aiml@dsu.edu.in

Ashwini Kodipalli

School of CSE (AI&ML), PES University, Bangalore, Karnataka, India ashwinik@pes.edu

Shridhar B Devamane

Dept. of CSE, MS Engineering College, Bangalore, Karnataka, India devamaneshridhar@gmail.com

Gargi N

Dept. of AI & Data Science, Global Academy of Technology, Bangalore, Karnataka, India gargi@gat.ac.in

Abstract.

Car or an automobile is a four-wheeled vehicle primarily designed for transportation on roads and has become an essential part of modern life. The automobile industry is considered one of the major indicators of a developed economy due to its contribution to industrial growth, mobility, and technological advancement. With over one billion cars currently in use worldwide, selecting the most suitable car has become a challenging task for consumers because of the availability of numerous models with varying features, costs, fuel efficiency, safety standards, and performance characteristics. Therefore, this paper presents a comprehensive analysis that integrates the Fuzzy TOPSIS and Analytic Hierarchy Process (AHP) approaches for effective car evaluation and selection. The proposed methodology considers multiple decision-making criteria. The obtained results demonstrate that the proposed hybrid approach can efficiently assist customers, automobile industries, and decision-makers in identifying the most suitable vehicle. Furthermore, this work supports the United Nations Sustainable Development Goals (SDGs), particularly SDG 9, 11, 12.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Keywords. Selecting a car brand, Fuzzy AHP, Fuzzy TOPSIS

1. INTRODUCTION

A French innovator Nicolas-joseph Cugnot developed a first steam powered car vehicle in 1769. Over the period additional features and controls have been added and made progressively more complex. The automative industry which is rapidly evolving. Where selecting a car becomes a complex decision making due to the overwhelming factors like performance, safety, reliability, brand reputation and individual preference. Traditional decision-making cannot be performed on multiple inheritance. To solve this problem applications of fuzzy logic developed. By using a fuzzy logic, we can obtain an accurate and tailored recommendation. This approach not only gives us decision making criteria but also allows us to prefer the priorities of individual consumers. This research introduces an efficient approach that combines the Fuzzy Technique for Order Preference by Similarity to an Ideal Solution (Fuzzy TOPSIS) with the Analytic Hierarchy Process (AHP).

2. RELATED WORK

[1] The authors have employed a conceptual framework to examine business dynamics and competitiveness through a hybrid integrated enterprise model. They utilized the FAHP ranking technique to identify solutions closest to the ideal. The hybrid approach, known as Fuzzy AHP-TOPSIS [2], was used to evaluate various aspects of web application information security, aiming to provide valuable and practical insights. [3] Eco-friendly cars are anticipated to become a major market force in global transportation and technology, as they are crucial for achieving environmental and societal sustainability. [4] This paper aims to analyze consumer perspectives on purchasing eco-friendly vehicles and assess the necessity of these vehicles. Incorporating FAHP can improve decision-making regarding the selection of environmentally friendly cars. The results from FAHP [5] are then compared with those obtained using the traditional Analytic Hierarchy Process (AHP) [6]. Four main water sources are assessed based on five criteria. The evaluation identified groundwater wells as the most suitable option. The findings provide valuable insights to help decision-makers and researchers make practical and effective decisions regarding the allocation of rural water resources. Paper [9] showed the analysis of mental health using fuzzy techniques such as AHP and TOPSIS. Password security strength analysis and factors influencing them was implemented in [10]. Prediction of Schizophrenia disease was carried out using the very popular methods such as Fuzzy AHP and TOPSIS methods in paper [11].

3. METHODOLOGY

Our research employs the Fuzzy techniques like fuzzy AHP, fuzzy AHP using a geometric means and fuzzy TOPSIS method in selecting the car brand based on multiple features. We define eight essential attributes: price (c1), Range kmph (c2), Capacity (c3), Seats (c4), Variants (c5), comfort (c6), and fuel type(c7). The output of this is giving a rank to the important attributes that will help in decision making of selection of a car brand. By knowing the relevant features, it is easy to remove the irrelevant or redundant features thereby reducing the computational expenses like power, memory and so on.

3.1. Fuzzy AHP

AHP is an effective method for calculating the weights of criteria related to a defined output. The process involves seven steps: Identify the problem and define the criteria; Evaluate the alternatives. The evaluation criteria consist of eight attributes used for selecting a car brand. Out of eight criteria we check which of them are really important for deciding car brand.

3.2. AHP using Geometric Mean

The technique of integrating qualitative and quantitative methods is known as the AHP process. AHP is mainly utilizes decision-making tool. It uses pair-wise comparisons for multiple criteria. This method consists of eight steps. Fuzzy AHP relative importance scale, which is used to express pairwise comparisons under uncertainty. Each importance level is shown using a Triangular Fuzzy Number (TFN) consisting of lower, middle, and upper values. A value of 1 refers to the condition that two criteria are equally important and is represented as (1,1,1). Values such as 3, 5, 7, and 9 show moderate, strong, very strong, and extreme importance respectively. For Intermediate judgments values such as 2, 4, 6, and 8 are being used when the preference lies between two adjacent importance levels. For example, strong importance is represented by (4,5,6). The reciprocal values 1/3, 1/5, 1/7, and 1/9 denote inverse comparisons when one criterion is less important than another. Using TFNs allows decision-makers to capture vagueness and subjectivity more effectively than exact numerical values.

3.3. TOPSIS

Triangular Fuzzy Numbers (TFNs) are used to convert linguistic evaluations into numerical fuzzy values. Terms such as Very Poor, Poor, Good, Very Good, and Extreme Good are expressed as fuzzy ranges to capture uncertainty and human judgment in decision-making.

4. RESULTS

4.1. AHP Method

The attribute with the highest criteria weight, 0.3587 (Price), is identified as the primary selecting factor, while the attribute with the lowest criteria weight, 0.0393 (Fuel Type), is determined to be the least influential factor in the selection process.

4.2. Fuzzy AHP

The normalized weights show the relative importance of each criterion, with Price (0.35) having the highest priority, followed by Seats (0.20) and Range (0.17). Criteria such as CC (0.10), Comfort (0.08), Variants (0.036), and Fuel Type (0.036) have lower influence on the final decision. Because the eigen values calculated are same as the ones obtained when AHP method was used, the pairwise comparison matrix is considered as consistent. This shows that the feature weights calculated is trustworthy, reliable and also stable. Also this can help us in building a hybrid technique such as concatenating two techniques, stacking two techniques which can be called as assembling two techniques. This can be used as future scope of this research paper.

4.3. Fuzzy TOPSIS

In this method, Alternative A7, with a closeness coefficient of 0.7585, ranks first. It is the most preferred choice among customers selecting a car, while Alternative A3, with a coefficient of 0.4967, is the least preferred.

5. CONCLUSION

In this paper we have used the AHP, Fuzzy AHP, and TOPSIS techniques for selecting the prevalent features for selecting a car brand. In the AHP method we defined the problem statement, written the hierarchy, defined the criteria, found the criteria weights, and then determined the rank for all the alternatives. The outputs got from AHP method proved to be similar to other MCDM methods. The second method that we used is the Fuzzy AHP method where we converted the crisp values into actual fuzzy numbers. When compared the AHP and Fuzzy AHP methods, the car selected via the Fuzzy AHP method was more correct and precise. The third method that we employed was the TOPSIS method where the Positive Ideal and Negative Ideal Solutions were considered and their distance were estimated from the alternatives. The closer the distance the better the decision was. After combining all three methods, the best criteria and the alternative was determined which helped in car purchase. When the ranking differed from one method to the other, Fuzzy AHP method was preferred.

6. REFERENCES

[1] D. Jun Yi Tey et al.: Novel NDAHP for MCDM Method: Case Study in KLSE A.

[2] Agrawal et al.: Evaluating the Security Impact of Healthcare Web Applications Through Fuzzy-Based Hybrid Approach

[3] W. Liang et al.: Performance Evaluation of Green Mine Using a Combined MCDM Method With Picture Fuzzy Information.

[4] N. V. Thanh et al.: Strategic Decision Making at a Steel Industry Assisted by Fuzzy Theory

[5] Adam Shariff Adli Aminuddin et al 2019 J. Phys.: Conf. Prediction of Schizophrenia in Patients Using Fuzzy AHP and TOPSIS

[6] Integrated Fuzzy TOPSIS And AHP Approach for Efficient Water Resources Allocation in Rural Areas

[7] Analysis of Mental Health of a Young Adult: Post Covid Using Fuzzy TOPSIS & Fuzzy AHP

[8] Enhancing Password Security: Fuzzy AHP and TOPSIS-based Vulnerability Assessment.

[9] Uttarkar, C., Arun, C. B., Anusha, M., Khushi, N., Kodipalli, A., & Rao, T. (2023, October). Analysis of Mental Health of a Young Adult: Post Covid Using Fuzzy TOPSIS & Fuzzy AHP. In 2023 4th IEEE Global Conference for Advancement in Technology (GCAT) (pp. 1-7). IEEE.

[10] Kumar, A., Kodipall, A., Rao, T., Soma, B., & Gargi, N. (2024, April). Enhancing Password Security: Fuzzy AHP and TOPSIS-based Vulnerability Assessment. In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

[11] Anoop, R., Anand, I., Rehan, M., Yashvanth, R., Kodipalli, A., Rao, T., & Kamal, S. (2023, March). Prediction of Schizophrenia in Patients Using Fuzzy AHP and TOPSIS Methods. In International Conference on Machine Learning, IoT and Big Data (pp. 517-527). Singapore: Springer Nature Singapore.

<IC-DAIS>
<Conference Series name> <Volume number and Year> <DOI Number>

Evaluating Water Potability: A Fuzzy Logic Approach

Trupthi Rao

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India trupthirao-aiml@dsu.edu.in

Shivnandan Rai

Dept. of CSE (DS), Dayananda Sagar University, Bangalore, Karnataka, India souramita-ds@dsu.edu.in

Souramita Bhowmik

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India shivnandanrai-aiml@dsu.edu.in

Ashwini Kodipalli

School of CSE (AI&ML), PES University, Bangalore, Karnataka, India ashwinik@pes.edu

Shridhar B Devamane

Dept. of CSE, MS Engineering College, Bangalore, Karnataka, India devamaneshridhar@gmail.com

Gargi N

Dept. of AI &DS, Global Academy of Technology, Bangalore, Karnataka, India gargi@gat.ac.in

Abstract.

Water is among the fundamental resources available on Earth, crucial in ensuring life, especially acting as drinking water. On the contrary, the resource distribution of water on Earth varies from one region to the other. Whereas some regions receive ample water supplies, others face water scarcity. This shows unequal potential and, therefore, demands regional analysis. This research determines whether water is potable or not based on the analysis of a dataset found on Kaggle1; specifically, the sources of water analysed here are from India. In this paper, we will apply the fuzzy AHP method to evaluate and prioritize various factors affecting the water potability. Then, by considering the obtained factors in priority order, we will develop a model for predicting the potability of water resources. As a result, water quality features are independent from each other, and all of them must satisfy the standard conditions so that water is considered potable. The present study offers some clear insight into the aspects of water quality assessment and a more feasible way of prediction concerning water potability based on given certain measurable factors.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Keywords. Fuzzy TOPSIS Water Probability, Fuzzy AHP, Prediction

1. INTRODUCTION

Rapid industrialization and urbanization in India have resulted in severe deterioration of water quality. The untreated sewage, rampant disposal of plastics and other pollutants into rivers, lakes, and ponds, along with industrial waste, has contaminated nearly 70% of the available water resources of the country. This has consequently emerged as an unprecedented water pollution crisis in which as many as 80% of the rural population and 20% of the urban population have no access to potable water. The implications are extremely serious: diarrhea, dysentery, typhoid, cholera, and hepatitis are waterborne diseases in the Literally millions of people get afflicted by these diseases every year. These diseases strike children with particular vengeance and, together, give some of the worst mortality rates, creating a vicious circle of poverty and ill health. In the backdrop of such facts, there is an immediate need for effective water quality monitoring to help overcome some of the challenges facing local communities. This paper presents the way fuzzy logic can help in solving the issues related to traditional methods for water quality assessment, especially when resources are limited. The awareness that this research possesses of the critical issue at hand-water potability-allows it to play its part in the worldwide attempt to ensure safe, accessible drinking water.

2. RELATED WORK

Over the years, various methods have been developed to assess water quality, with traditional approaches often relying on statistical techniques to calculate water quality indices (WQIs) [1]. Fuzzy logic was first introduced by Lotfi Zadeh in the 1960s as a mathematical framework for dealing with imprecision and reasoning that mimic human decision-making processes. [2] Unlike traditional binary logic, which categorizes data into discrete classes, fuzzy logic allows for degrees of membership in different sets, making it particularly suitable for environmental applications where data often contains ambiguities and uncertainties. Fuzzy logic offers a more flexible and precise approach to water quality assessment. [3] By incorporating fuzzy reasoning, a fuzzy logic-based WQI can map input determinants, such as physicochemical parameters, to output results that better reflect the continuous nature of water quality data. In the fuzzification stage, real-world data is transformed into fuzzy sets using membership functions, which represent the degree to which each data point belongs to a specific set. [4] For example, instead of classifying water as simply "good" or "bad," fuzzy logic allows for a range of values that can more accurately describe the water's quality. [5] Fuzzy set operations are then used to derive fuzzy results, which are subsequently converted into real values through the process of defuzzification. [6] This is particularly important in water quality management, where factors such as pollution levels, seasonal changes, and human activities can cause significant fluctuations in water quality. [7] This flexibility makes fuzzy logic an ideal tool for assessing water quality in diverse environments, from urban areas with complex pollution sources to rural regions where data may be sparse or incomplete. Additionally, fuzzy logic can incorporate additional critical parameters that may not be included in traditional WQIs, such as heavy metal content or emerging contaminants, further enhancing the accuracy of the assessment [8]. Bellur et al. [9] assessed water quality using ML techniques.

3. METHODOLOGY

There are many methods to choose the best parameters to minimize the water loss. But we are going to consider only some methods for the evaluation.

3.1. AHP

The Analytical Hierarchy Process (AHP) is a multi-criteria decision-making technique that aids in prioritizing and evaluating complex options by organizing them into a hierarchical structure. This method is commonly used by professionals to facilitate strategic planning and guide decisions effectively.

3.2. FAHP

Fuzzy AHP extends the traditional AHP by incorporating fuzzy logic to better handle uncertainties and ambiguities in decision-making. Fuzzy AHP is particularly useful when human judgments are imprecise or uncertain, and it offers a systematic way to address these complexities.

4. RESULTS

4.1. AHP

In this analysis, we evaluated the potability of water samples based on several criteria. The consistency of our judgment matrix was confirmed with a Consistency Index value of 0.05, which aligns well with the random consistency index of 0.0556. This indicates that our judgments regarding the criteria are consistent and reliable. These ranking highlights that the most significant factors in determining water potability are pH and Hardness, followed by Solids, Conductivity, and Organic Carbon. The criteria Chloramines, Sulphate, Trihalomethanes, and Turbidity have less impact on the final assessment. The AHP analysis identified pH as the most important water quality parameter (0.405), followed by Hardness (0.247) and Solids (0.221).

4.2. Fuzzy AHP

In the final step the normalized values of all attributes are obtained which is seen in Table I. The fuzzy AHP analysis provides a robust framework for assessing water potability by accommodating the uncertainty and vagueness inherent in evaluating multiple criteria. The findings highlight the importance of Trihalomethanes and Conductivity in determining water quality, suggesting that these factors should be prioritized in water quality assessments and treatment processes. By applying fuzzy AHP, we can more effectively prioritize and manage the criteria affecting water potability, leading to better-informed decisions and improved water safety. This method helps in making nuanced assessments that account for the complexities and uncertainties involved in evaluating water quality.

TABLE I: Normalized Data

Sl.no	Normalized values	Condition
1	0.1463	ph
2	0.126	Hardness
3	0.5937	Solids
4	0.2766	Conductivity
5	0.5342	Organic carbon
6	0.6468	Chloramines
7	0.2622	Sulphate
8	0.7242	Trihalomethane
9	0.0023	Turbidity

5. CONCLUSION

In conclusion, the integration of fuzzy logic with AHP offers a robust framework for water quality assessment, accommodating the inherent uncertainties and complexities of evaluating multiple criteria. This approach enables more informed decision-making and enhances water safety management. By prioritizing critical factors such as Trihalomethanes and Conductivity, policymakers and water resource managers can better address water quality issues and implement effective preventive measures. The advancements presented in this study have the potential to revolutionize water quality monitoring practices, contributing significantly to public health and sustainable water management globally.

6. REFERENCES

[1] Proposition of New Ensemble Data-Intelligence Models for Surface Water Quality Prediction IEEE Access (Volume: 9) DOI: 10.1109/ACCESS.2021.3100490.

[2] Quality Risk Analysis for Sustainable Smart Water Supply Using Data Perception IEEE Transactions on Sustainable Computing (Volume: 5, Issue: 3, 01 July-Sept. 2020) Page(s): 377 – 388 DOI: 10.1109/TSUSC.2019.2929953.

[3] WaterNet: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes Published in: IEEE Access (Volume: 10).
Fuzzy Logic Water Quality Index and Importance of Water Quality Parameters Reinier Bouwmeesterand S Mohan DOI:https://doi.org/10.4137/ASWR.S2156.

[5] A novel approach in water quality assessment based on fuzzy logic Hamed Gharibi, Amir Hossein Mahvi, Ramin Nabizadeh.

[6] Fuzzy evaluation of water quality classification Au8thor: lYilmaz Icaga DOI:https://doi.org/10.1016/j.ecolind.2006.08.002

[7] Febriansyah, J., Gernowo, R. and Kusumawardhani, A., 2017. Implementation of AHP and TOPSIS Method to Determine the Priority of Improving the Management of Government's Assets. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 4(03), pp.46-53.

[8] Minatour, Y., Khazaei, J. and Gholami, A., 2013, January. Extension of an integrated AHP and TOPSIS approach in water resources management. In 9th International River Engineering Conference, Shahid Chamran University, Ah, Iran.

[9] Bellur, A. P., VL, V. K., Choudary, S., Kodipalli, A., Rao, T., & Pushpalatha, V. (2023, June). Water Quality Assessment Using Machine Learning: A Comparative Analysis. In 2023 International Conference on Computational Intelligence for Information, Security and Communication Applications (CIISCA) (pp. 320-325). IEEE.

TabChain-Ensemble: A Blockchain-Integrated Ensemble Learning Framework for Secure IIoT Transactions

G. Bharanidharan[1], Bhavana Jamalpur[2], Dinesh P. S[3], Madhumitha C[4], P. X. Edwin Arshina[5], B. Jegajothi[6]*

[1] Associate Professor, Department of Computer Applications, The New College, Chennai – 600014, Tamil Nadu, India
Email: bharanidharan@thenewcollege.edu.in

[2] Associate Professor, School of CS & AI, SR University, Warangal, Telangana, India
Email: j.bhavana@sru.edu.in

[3] Associate Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Alathukombai (Post), Sathyamangalam – 638401, Tamil Nadu, India
Email: dineshps@bitsathy.ac.in

[4] Research Scholar, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu – 600062, India
Email: madhuphd.cs24@gmail.com

[5] Assistant Professor, Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, Tamil Nadu, India
Email: arshna.se@vistas.ac.in

[6] Research Associate, SRS Tech Solutions, Chennai, Tamil Nadu, India
Email: jegajothibalakrishnan@gmail.com

***Corresponding author:** Dr. B. Jegajothi (jegajothibalakrishnan@gmail.com)

Abstract: The use of industrial Internet of Things (IIoT) environments is becoming more susceptible to cyber-attacks because of distributed and resource-constrained nature. Intelligent intrusion detection systems can only ensure secure and reliable process of transaction in such systems by incorporating secure data management systems. The present paper is a proposal of TabChain-Ensemble, a new machine learning-enhanced blockchain architecture of secure IIoT transactions. To achieve a high rate of detection and data integrity the proposed approach incorporates an ensemble learning model that delivers a combination of XGBoost, LightGBM, and Random Forest together with a permissioned blockchain layer. The publicly available TON IoT network dataset is used to assess the model; it includes realistic IIoT traffic labeled with attack and normal samples. Enhancement of detection performance is done using feature preprocessing, selection and optimization of the model. As experimental findings indicate, the suggested TabChain-Ensemble has the high accuracy of 99.55% and F1-score of 99.55, which is better than that of the single basic models. This model also has a ROC-AUC of 0.9998, which shows a great discriminative ability. Additionally, there is a small computational cost associated with integrating blockchain, whereby the latency cost rises by 20-25 % and the throughput decreases,

but the overall system security and auditability is also improved substantially. The findings ensure that the suggested framework is able to offer a good balance between detection performance and secure transaction management, and can be applicable to real-world IIoT applications.

Keywords: Industrial Internet of Things (IIoT), Intrusion Detection System (IDS), Ensemble Learning, XGBoost, LightGBM, Random Forest, Blockchain, Secure Transactions, TON_IoT Dataset, Cybersecurity, Machine Learning, Precision-Recall Analysis

1. INTRODUCTION

The fast development of the Industrial Internet of Things (IIoT) has made the flow of interconnection between industrial devices completely smooth, allowing to increase automation, efficiency, and decision-making based on data. Nevertheless, such enhanced interconnectivity also puts IIoT systems at risk of a large variety of cybersecurity threats such as data breaches, denial-of-service attacks, and unauthorized access. Conventional security controls are not usually able to manage the dynamic and distributed character of IIoT settings. Accordingly, machine learning-based intelligent intrusion detection systems (IDS) have received enormous interest as they are capable of identifying complex and changing attack patterns. Simultaneously, blockchain technology has become a hopeful future of providing a system of safe, transparent, and non-modifiable data transaction within a decentralized system. The TabChain-Ensemble framework proposes the combination of several machine learning models, namely XGBoost, LightGBM, and Random Forest to enhance the performance of detection based on ensemble learning. There is also a blockchain layer to provide a means of safe, transparent and tamper-resistant storage of transactions, which would increase the overall trustworthiness and safety of IIoT systems. The fast development of the Industrial Internet of Things (IIoT) has made the interconnection between industrial devices seamlessly, which makes it easier to monitor in real-time, automate, and make smart decisions. Nevertheless, it is true that the scale and heterogeneity of IIoT systems have been growing at a rapid pace, thus increasing its attack surface that makes it highly susceptible to advanced cyber threats. Secure communication, data integrity and trust management in these distributed environments is a critical issue that has to be ensured. These issues have been considered in recent research that accesses more advanced security mechanisms to include federated learning, blockchain, and multi-agent reinforcement learning, including quantum-secure and trust-aware IIoT systems [1].

The machine learning and deep learning methods have gained high usage in intrusion detection and anomaly detection tasks in IIoT settings because of the possibility to recognize intricate attack patterns. Lightweight deep learning hybrid models have been suggested in order to provide privacy-aware data processing especially in resource-constrained industrial and medical IoT systems [2]. Also, privacy-sensitive machine learning models are created to facilitate sharing of data and collaborative learning without exposing sensitive information in industrial networks [3]. Although this has been made possible, most machine learning-based solutions have some difficulties in areas of data centralization, scalability, and susceptibility to adversarial attacks.

To address these restrictions, machine learning has been applied to blockchain technology to offer decentralized, secure, and tamper-resistant data management. Blockchain-based systems with deep learning have been shown to be more secure to the IIoT systems through data integrity and secure communication among distributed nodes [4]. In addition, federated learning methods with blockchain have also been suggested to facilitate safe model aggregation and privacy protection within IoT settings [5]. On the same note, blockchain

and federated learning have been demonstrated to enhance data sharing that ensures privacy and yet does not compromise trust among the devices involved [6]. Some recent publications have also examined collaborative threat intelligence systems that combine blockchain with machine learning to boost the security of IoT by sharing knowledge on a distributed basis [7]. Moreover, federated structures of learning based on blockchain technology with smart contracts have been implemented to provide safe coordination and automation of IoT systems [8].

Regardless of these dramatic developments, the current solutions are usually characterized by such major flaws as hefty computational costs, inability to process tabular intrusion data with efficiency, and the absence of effective combination of detection metrics with system-level security indicators. Most of the frameworks are oriented towards better detection accuracy or better data security but lacks a balance approach that will offer high detection performance, scalability, and effective management of transactions. Furthermore, ensemble learning methods have not been considered sufficiently in blockchain-based architectures in terms of enhancing intrusion detection in IIoT systems.

In order to overcome these issues, this paper suggests TabChain-Ensemble, an ensemble learning technology that includes blockchain and is aimed at managing secure and efficient IIoT transactions. The suggested solution is a combination of various machine learning systems to detect the data with more accuracy and use blockchain technology to provide secure, transparent, and resistant data handling, which will be a full-fledged solution to current IIoT security needs.

2. RELATED SURVEY

The combination of machine learning and blockchain technologies to increase the security of IIoT systems has been studied in the recent research. Some of the research studies have employed machine learning algorithms, including the Random Forest, Support Vector Machines, and deep learning in intrusion detection, and have demonstrated better detection accuracy on datasets like TON_IoT and N-BaIoT. The XGBoost and LightGBM techniques have been shown to perform better in ensemble learning with high-dimensionality and tabular intrusion data. Also, blockchain-based architectures have been recommended to provide integrity of data, secure communications, and decentralized access control of IIoT systems. Nevertheless, a lot of solutions that are available only discuss detection performance and no consideration of system-level security or they add a lot of computational load without optimization of detection efficiency. This brings to light the necessity of developing a unified structure that would maintain high detection rates and provide secure and efficient offering of transactions in the distributed setting by ensuring the concerns of privacy, scalability, and trust are met. ecent developments of Industrial Internet of Things (IIoT) security have been largely driven by the combined efforts of blockchain and federated learning (FL) to counter the issues of privacy, scalability and trust in the distributed setting. Salim et al. [8] created a federated learning model with block chains and smart contracts to provide security in the coordination and data integrity of the IoT systems. On the same note, Kumar et al. [9] proposed a privacy-saving and secure framework based on machine learning on blockchain-based smart city applications, which show better protection of data and resilience of the system.

Rahman et al. [10] also extended this paradigm to suggest a decentralized approach of federated learning, that is blockchain controlled, to improve data provenance and security in the Internet of Health Things (IoHT) setting. Meanwhile, Kumar and Tripathi [11] proposed a profound blockchain-based trustful and privacy-aware IIoT systems, which is concerned with the secure information exchange and establishment of trust. The article by

Jamil et al. [12] also examines the combination of blockchain and machine learning in IoT-enabled systems, and presents a safe framework of using the system in keeping fitness applications that guarantees the safety and dependability of data.
In an attempt to enhance the efficiency of data sharing in industrial sectors, Kumar et al. [13] suggested a permissioned blockchain with deep learning to facilitate efficient exchange of data in industrial healthcare systems with security. Furthermore, Ali et al. [14] provided an extensive overview of blockchain-based intrusion detection methods and federated learning-based intrusion detection methods in edge-enabled IIoT networks, in addition to emphasizing the advantages of the decentralized system of security, as well as revealing the drawbacks of computational load and scalability.
More research has explored federated learning in cybersecurity in IoT setups. Ferrag et al. [15] conducted a detailed review of the federated deep learning methods and proved their usefulness in distributed intrusion detection and collaborative threat intelligence. Based on this idea, Yazdinejad et al. [16] developed a federated learning-based cyber threat hunting system that is combined with blockchain, which allows the detection of cyber threats in IIoT networks in a scalable and secure manner. Although the capability of blockchain and federated learning integration on IIoT security has greatly improved, the current practices mainly focus on privacy protection and decentralization of trust and neglect effective and high-performance intrusion-detection systems. A large number of frameworks are based on methods of single-model learning, which could restrict the detection accuracy and robustness when working with complex and high-dimensional tabular data like TON_IoT. What is more, not much attention has been paid to ensemble learning approaches in blockchain-integrated architectures, and to the assessment of system-level performance metrics, such as latency, and throughput. The above limitations point to the necessity of a framework that would be integrative, in both accuracy of detection and safe and effective management of transactions, which is the driving force behind the proposal of the TabChain-Ensemble approach.

3. PROPOSED METHODOLOGY

The suggested TabChain-Ensemble model is aimed at offering a comprehensive solution to the problem of efficient and safe intrusion detection in the Industrial Internet of Things (IIoT) via ensemble machine learning and blockchain-based transaction control. The framework employs a data-based method where the traffic of an IIoT network is preprocessed with normalization, encoding and feature selection methods, to guarantee input representation of high quality. This is followed by an ensemble learning model, which consists of XGBoost, LightGBM, and the Random Forest, to boost the detection performance based on the complementary advantage of each of the classifiers, to improve the robustness and generalization to different patterns of attacks. The ensemble decision scheme combines the predictions based on soft voting scheme which allows the probabilistic integration of the model results to enhance the classification rates. Simultaneously, a permissioned blockchain layer is introduced to have secure, transparent, and tamper-resistant transaction processing, with the result outcome of every detection being cryptographically hashed and stored as a validated block, and thus, no malicious alteration can be made to the data, and the integrity of the data is guaranteed. The combination of machine learning and blockchain blocks will enable the framework to perform intrusion detection and ensure data security at the same time, as well as system-level assessment with the help of the analysis of latency and throughput. This is a holistic design that may guarantee the proposed framework not only a high detection accuracy of the tabular datasets like TON_IoT, but also reliability and trustworthiness of distributed IIoT systems. Fig. 1

depicts the overall flow of Proposed TabChain-Ensemble architecture for secure IIoT intrusion detection and blockchain-based transaction management.

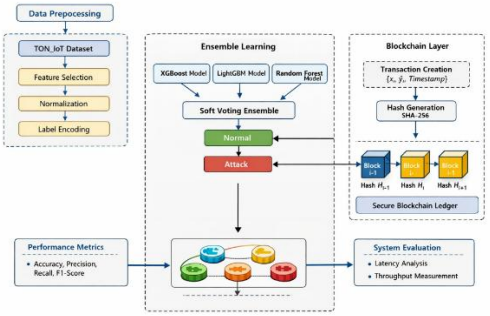


Fig. 1. Proposed TabChain-Ensemble architecture for secure IIoT intrusion detection and blockchain-based transaction management.

3.1 Dataset details

The case study is based on the publicly accessible TONIoT (Telemetry, Operating system, and Network IoT) dataset that is knownly created with a purpose to study cybersecurity in Industrial Internet of Things (IIoT) settings. The data is obtained based on a realistic cyber range testbed which emulates industrial IoT networks and includes heterogeneous devices, network traffic, and attack scenarios. It has an extensive list of telemetry and network flow capabilities that log both normal and malicious traffic of various types of attacks. The dataset of the TONIoT network utilized in this study is a structured tabular dataset comprising of a set of numerous numerical attributes that are a representation of both packet-level and flow-level attributes including the packets size, duration, protocol attributes, and statistical traffic characteristics [17]. The dataset is balanced with both benign and attack examples of the most typical cyber threats of Distributed Denial of Service (DDoS), ransomware, injection attacks, and reconnaissance operations. In order to have high-quality input to machine learning models, data preprocessing processes such as label encoding, data normalization, elimination of redundant identifiers (e.g., timestamps and IP addresses), and feature selection are conducted before model training. Stratified sampling is then applied to the dataset to divide it into training and testing set to maintain the distribution of the classes. The TON_IoT dataset is most commonly known as a benchmark dataset because of its realistic design and labeled attack scenarios to gain knowledge of using intrusion detection systems in IIoT settings. Table I shows the Dataset summary.

Table I: Dataset summary

Attribute	**Description**
Dataset Name	TON_IoT (Telemetry, Operating system, and Network IoT)
Data Type	Tabular (Network Traffic Data)
Total Samples	~82,000 (approx., based on subset used)
Training Samples	~65,000
Testing Samples	~16,000
Number of Features	16 (after preprocessing)

Feature Type	Numerical (after encoding & cleaning)
Classes	Binary (Normal, Attack)
Attack Types	DDoS, DoS, Ransomware, Injection, Reconnaissance
Data Source	UNSW Canberra Cyber Range Lab
Preprocessing Steps	Label Encoding, Feature Selection, Normalization, Removal of identifiers
Train-Test Split	80:20 (Stratified)
Application	Intrusion Detection in IIoT

3.2 Mathematical Modeling and Analysis

The objective of the proposed TabChain-Ensemble framework is to design a secure and efficient intrusion detection mechanism for Industrial Internet of Things (IIoT) environments by integrating ensemble machine learning with blockchain-based validation. Let the input dataset be denoted as $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$represents the feature vector of the i-th network instance with $d$attributes, and $y_i \in \{0,1\}$denotes the corresponding class label, where 0 represents normal traffic and 1 represents malicious activity [18].

The proposed framework employs an ensemble learning model consisting of multiple base classifiers, namely XGBoost, LightGBM, and Random Forest. Let the prediction function of each model be defined as $f_k(x_i)$, where $k \in \{1,2,3\}$corresponds to the individual classifiers. The ensemble prediction is obtained using a soft voting mechanism, where the final predicted probability is computed as the average of individual model probabilities:

$$P(y = 1 \mid x_i) = \frac{1}{K}\sum_{k=1}^{K} f_k\,(x_i) \qquad (1)$$

where Kdenotes the total number of models in the ensemble. The final class label is then determined using a decision threshold τ, defined as:

$$\hat{y}_i = \begin{cases} 1, & \text{if } P(y = 1 \mid x_i) \geq \tau \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

To evaluate the classification performance, standard metrics are defined based on the confusion matrix components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The accuracy, precision, recall, and F1-score are computed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (5)$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (6)$$

The Receiver Operating Characteristic (ROC) curve is constructed by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR), defined as:

$$\text{TPR} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN} \qquad (7)$$

In addition to detection performance, the proposed framework incorporates a blockchain layer to ensure secure transaction management. Each prediction result is converted into a transaction and stored in the blockchain using a cryptographic hash function. Let T_i denote the transaction corresponding to the i-th instance. The block hash is computed as:

$$H_i = \text{SHA-256}(T_i \parallel H_{i-1}) \qquad (8)$$

where H_{i-1} represents the hash of the previous block, and $\parallel$ denotes concatenation. This chaining mechanism ensures immutability and tamper resistance of stored transactions.
The system performance is further analyzed using latency and throughput metrics. Latency is defined as the time required to process a batch of transactions:

$$\text{Latency} = t_{\text{end}} - t_{\text{start}} \qquad (9)$$

where t_{start} and t_{end} denote the start and end times of processing. Throughput is defined as the number of processed instances per unit time:

$$\text{Throughput} = \frac{N}{\text{Latency}} \qquad (10)$$

The proposed formulation ensures that the TabChain-Ensemble framework achieves high classification performance while maintaining secure and efficient transaction processing. The integration of ensemble learning improves predictive robustness, whereas the blockchain layer guarantees data integrity and trust, making the system suitable for real-world IIoT deployments [19]. The mathematical formulation of the TabChain-Ensemble framework establishes a structured relationship between ensemble-based prediction and blockchain-driven validation, enabling a comprehensive analysis of both detection performance and system-level security. The ensemble model enhances classification reliability by minimizing variance and bias through the aggregation of multiple learners, while the blockchain layer ensures immutable and verifiable storage of prediction outcomes using cryptographic hashing mechanisms. This dual-layer integration enables the framework to address critical IIoT challenges, including accurate intrusion detection, data integrity, and secure transaction management. Furthermore, the proposed approach maintains a balance between computational efficiency and security overhead, as demonstrated through latency and throughput evaluation. By combining high-performance ensemble learning with decentralized blockchain validation, the framework provides a scalable and reliable solution suitable for real-world IIoT deployments, where both detection accuracy and system trust are essential [20].

Algorithm 1: TabChain-Ensemble Framework for Secure IIoT Intrusion Detection
Input: TON_IoT dataset *D* **Output:** Predicted labels and blockchain-secured transactions
Load the TON_IoT dataset *D* Perform data preprocessing Remove irrelevant attributes (ID, timestamp, IP addresses) Encode class labels (Normal = 0, Attack = 1) Normalize feature values

Apply feature selection
Split the dataset into training and testing sets
Initialize machine learning models
XGBoost model
LightGBM model
Random Forest model
Train all models using the training dataset
For each test instance x_i:
a. Obtain prediction probabilities from each model
b. Compute the average probability (ensemble output)
c. Assign final label based on threshold τ
d. Create a transaction containing prediction details
e. Generate hash using SHA-256
f. Store the transaction in the blockchain
Evaluate model performance and Measure Latency, Throughput

4. RESULTS AND DISCUSSION

The TON IoT dataset was used to assess the performance of the proposed framework of TabChain-Ensemble in a controlled experimental environment. The dataset was preprocessed by using features selection, normalization and removal of irrelevant attributes after which 80:20 stratified train-test split was done to maintain the distribution of classes. The suggested model, which is a combination of the XGBoost, LightGBM, and Random Forest were compared to the independent models which serve as the baseline models. Standard classification measures, including accuracy, precision, recall, F1-score, ROC-AUC, and specificity, were used to measure performance, and system level measures, such as latency and throughput were also used.

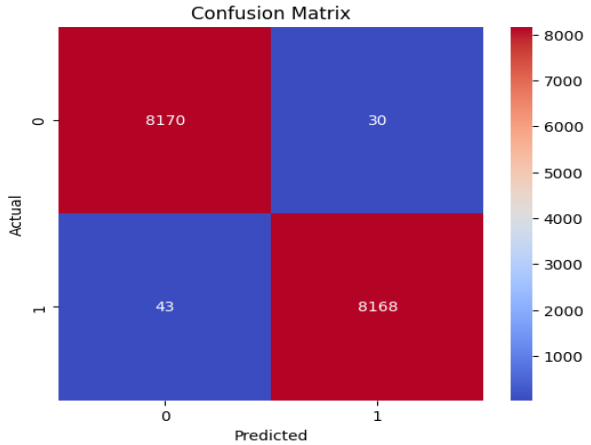


Fig. 2. Confusion matrix of the proposed TabChain-Ensemble model showing classification performance for normal and attack instances.

In Fig. 2, the confusion matrix of the suggested TabChain-Ensemble model is provided. One will note that the model has a high number of true positives and true negatives whilst having very low false positives and false negatives. This shows that the proposed framework is very effective in the accurate identification of normal and attack cases, which is important in the intrusion detection in IIoT systems. The even spread of the classification results also affirms that the model does not have any bias towards a particular class.

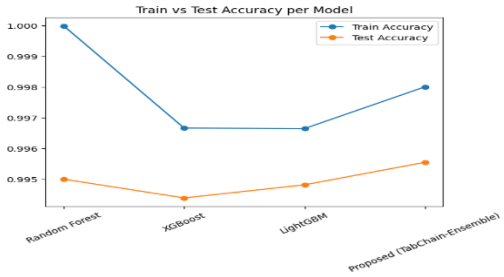


Fig. 3. Training and testing accuracy comparison of baseline models and the proposed TabChain-Ensemble framework.

The comparison of the train and test accuracy of all models is presented in Fig. 3. These findings indicate that there exists a small difference between training and testing accuracy, which means that the models, specifically the proposed ensemble, have high generalization with insignificant overfitting. Although the training accuracy of Random Forest is a bit higher, its slight decrease in test accuracy indicates that it is slightly overfitting, where the proposed model is more balanced with bias and variance.

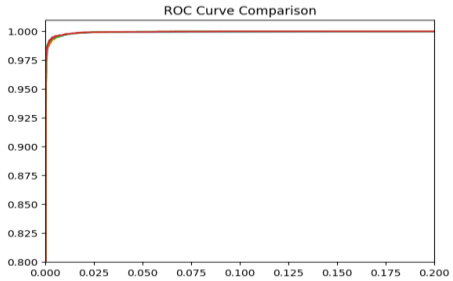


Fig. 4. Receiver Operating Characteristic (ROC) curve comparison of different models, highlighting classification performance in the low false-positive region.

The Receiver Operating Characteristic (ROC) curve is illustrated in Fig. 4 and the low false-positive area is zoomed in. All models have high ROC-AUC values (ranging between 0.999), which means that they have a high discriminative ability. Nevertheless, the proposed TabChain-Ensemble has a slight lead in the critical low false-positive area which is especially significant in IIoT security systems when the false alarms should be reduced to a minimum.

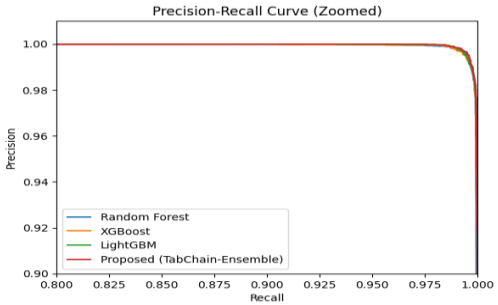


Fig. 5. Precision–Recall (PR) curve comparison illustrating model performance under varying recall levels.

Fig.5 demonstrates Precision-Recall (PR) curve, as well as it is zoomed to emphasize minor differences. The developed model has a higher precision with high recall values than the

baseline models do, meaning that it is very strong in identifying intrusion instances, without highly inflating the false positives. This is particularly applicable in skewed or security sensitive data in which precision-recall is more informative than accuracy

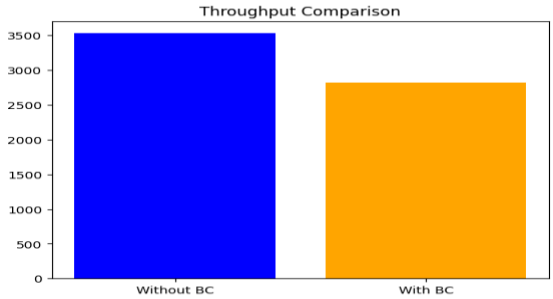


Fig. 6. Throughput comparison of the system with and without blockchain integration.

Fig. 6 depict the system-level analysis of performance of throughput with and without blockchain integration. The findings have shown that the addition of the blockchain layer has presented a median rise in latency (around 20-25 percent) as well as a compensatory reduction in throughput. Nonetheless, the trade-off can be tolerated due to the increased security, transparency, and data integrity offered by the blockchain mechanism, which is why it can be used in secure IIoT deployments. Table II presents the overall quantitative analysis of the performance of all models. As can be verified, the TabChain-Ensemble, with its accuracy (0.9955), F1-score (0.9955), and specificity (0.9962), is the most appropriate and has the best performance. The consistency of the results of the precision and recall values shows that the model is good in balancing false positive and false negative. There is also an indication that the ROC-AUC values in all the models are high, which indicates that the TON IoT data set is highly separable, but the proposed model is persistently doing better than the rest, which confirms its efficiency.

Table II: Classification Performance Comparison

Model	**Accuracy (ACC)**	**Precision**	**Recall (SEN)**	**F1-Score**	**ROC-AUC**	**Specificity (SPE)**
Random Forest	0.9949	0.9954	0.9945	0.9949	0.9996	0.9954
XGBoost	0.9944	0.9942	0.9946	0.9944	0.9998	0.9941
LightGBM	0.9948	0.9959	0.9938	0.9948	0.9998	0.9959
Proposed (TabChain-Ensemble)	**0.9955**	**0.9962**	**0.9948**	**0.9955**	**0.9998**	**0.9962**

Fig. 7 shows the increased performance metrics comparison of all the models evaluated in groups. The TabChain-Ensemble is always ranked with the best score in all the measures, such as accuracy (99.55%), precision (99.62%), recall (99.47%), and F1-score (99.55%). The enhancements over the individual models (XGBoost and LightGBM) are not very high, but they remain stable throughout all the evaluation criteria, which proves the usefulness of ensemble learning in improving the robustness and generalization of a model. Comparatively, the proposed framework is better performing compared to the traditional machine learning models and it is consistent with the existing studies that incorporate machine learning and blockchain to provide security to the IIoT. As opposed to the current methods which are mainly concerned with either accuracy in detection or security process, TabChain-Ensemble is an all round solution since it offers high-performance intrusion detection as well as secure transaction management. Enhanced predictive performance is

achieved by the integration of ensemble learning, and trust and data integrity are achieved by the blockchain layer. All these attributes render the proposed TabChain-Ensemble framework a useful and scalable solution to the real-world IIoT security application.

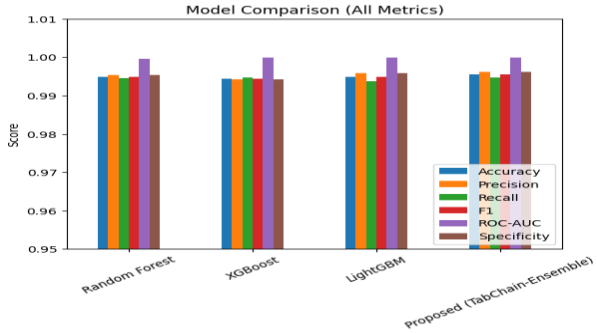


Fig. 7. Comparison of classification performance metrics (accuracy, precision, recall, F1-score, ROC-AUC, and specificity) across different models.

5. CONCLUSION

A new blockchain-enhanced ensemble learning framework presented in this paper includes TabChain-Ensemble, which is an ensemble learning framework developed to guarantee safe and precise intrusion detection in the context of the Industrial Internet of Things (IIoT). The presented solution is based on the XGBoost, LightGBM, and the Random Forest strengths to enhance the classification robustness and generalization, and introduce a permissioned blockchain layer to guarantee a secure, transparent, and tamper-resistant process of managing transactions. Evaluating the TON_IoT dataset experimentally, it was revealed that the proposed model has better performance with the highest accuracy and F1-score of 99.55% and a high ROC-AUC of 0.9998 compared to single base models. The findings also verify that the ensemble method is an efficient way to minimize the classification error, and the blockchain implementation can increase the data integrity at a reasonable cost of latency and throughput. These results demonstrate the practicality of the suggested framework in terms of maintaining the accuracy of detection and system-wide security, and it is applicable to the IIoT in the field. Moreover, the incorporation of lightweight blockchain software and edge-based federated learning may also be added to minimize the computational load and enhance the scalability. The study of adaptive ensemble strategies and real-time implementation in dynamic IIoT environments is also an interesting avenue to develop a better system responsiveness and robustness. In general, the suggested TabChain-Ensemble framework can be a powerful basis of intelligent, scalable and secure IIoT security solutions.

REFERENCES

1. Alandjani, G. (2025). A MARL-federated blockchain-based quantum secure framework for trust management in industrial internet of things. *Scientific Reports*, 15. https://doi.org/10.1038/s41598-025-23055-2
2. Almaiah, M., Ali, A., Hajjej, F., Pasha, M., & Alohali, M. (2022). A Lightweight Hybrid Deep Learning Privacy Preserving Model for FC-Based Industrial Internet of Medical Things. *Sensors (Basel, Switzerland)*, 22. https://doi.org/10.3390/s22062112
3. Arachchige, P., Bertók, P., Khalil, I., Liu, D., Çamtepe, S., & Atiquzzaman, M. (2020). A Trustworthy Privacy Preserving Framework for Machine Learning in Industrial IoT Systems. *IEEE Transactions on Industrial Informatics*, 16, 6092-6102. https://doi.org/10.1109/tii.2020.2974555
4. Aljuhani, A., Kumar, P., Alanazi, R., Albalawi, T., Taouali, O., Islam, A., Kumar, N., & Alazab, M. (2024). A Deep-Learning-Integrated Blockchain Framework for Securing Industrial IoT. *IEEE Internet of Things Journal*, 11, 7817-7827. https://doi.org/10.1109/jiot.2023.3316669

5. Kalapaaking, A., Khalil, I., Rahman, M., Atiquzzaman, M., Yi, X., & Almashor, M. (2023). Blockchain-Based Federated Learning With Secure Aggregation in Trusted Execution Environment for Internet-of-Things. *IEEE Transactions on Industrial Informatics*, 19, 1703-1714. https://doi.org/10.1109/tii.2022.3170348
6. Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT. *IEEE Transactions on Industrial Informatics*, 16, 4177-4186. https://doi.org/10.1109/tii.2019.2942190
7. Nazir, A., He, J., Zhu, N., Wajahat, A., Ullah, F., Qureshi, S., , X., & Pathan, M. (2024). Collaborative threat intelligence: Enhancing IoT security through blockchain and machine learning integration. *J. King Saud Univ. Comput. Inf. Sci.*, 36, 101939. https://doi.org/10.1016/j.jksuci.2024.101939
8. Salim, S., Moustafa, N., & Turnbull, B. (2025). BFL-SC: A blockchain-enabled federated learning framework, with smart contracts, for securing social media-integrated internet of things systems. *Ad Hoc Networks*, 169, 103760. https://doi.org/10.1016/j.adhoc.2025.103760
9. Kumar, P., Kumar, R., Srivastava, G., Gupta, G., Tripathi, R., Gadekallu, T., & Xiong, N. (2021). PPSF: A Privacy-Preserving and Secure Framework Using Blockchain-Based Machine-Learning for IoT-Driven Smart Cities. *IEEE Transactions on Network Science and Engineering*, 8, 2326-2341. https://doi.org/10.1109/tnse.2021.3089435
10. Rahman, M., Hossain, M., Islam, M., Alrajeh, N., & Muhammad, G. (2020). Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach. *Ieee Access*, 8, 205071 - 205087. https://doi.org/10.1109/access.2020.3037474
11. Kumar, R., & Tripathi, R. (2021). DBTP2SF: A deep blockchain-based trustworthy privacy-preserving secured framework in industrial internet of things systems. *Transactions on Emerging Telecommunications Technologies*, 32. https://doi.org/10.1002/ett.4222
12. Jamil, F., Kahng, H., Kim, S., & Kim, D. (2021). Towards Secure Fitness Framework Based on IoT-Enabled Blockchain Network Integrated with Machine Learning Algorithms. *Sensors (Basel, Switzerland)*, 21. https://doi.org/10.3390/s21051640
13. Kumar, R., Kumar, P., Tripathi, R., Gupta, G., Islam, A., & Shorfuzzaman, M. (2022). Permissioned Blockchain and Deep Learning for Secure and Efficient Data Sharing in Industrial Healthcare Systems. *IEEE Transactions on Industrial Informatics*, 18, 8065-8073. https://doi.org/10.1109/tii.2022.3161631
14. Ali, S., Li, Q., & Yousafzai, A. (2023). Blockchain and federated learning-based intrusion detection approaches for edge-enabled industrial IoT networks: a survey. *Ad Hoc Networks*, 152, 103320. https://doi.org/10.1016/j.adhoc.2023.103320
15. Ferrag, M., Friha, O., Maglaras, L., & Janicke, H. (2021). Federated Deep Learning for Cyber Security in the Internet of Things: Concepts, Applications, and Experimental Analysis. *IEEE Access*, PP, 1-1. https://doi.org/10.1109/access.2021.3118642
16. Yazdinejad, A., Dehghantanha, A., Parizi, R., Hammoudeh, M., Karimipour, H., & Srivastava, G. (2022). Block Hunter: Federated Learning for Cyber Threat Hunting in Blockchain-Based IIoT Networks. *IEEE Transactions on Industrial Informatics*, 18, 8356-8366. https://doi.org/10.1109/tii.2022.3168011
17. Rokh. (2025). Subset of the TON_IoT Dataset for IoT Intrusion Detection [Data set]. Zenodo. https://doi.org/10.5281/zenodo.18074069
18. Hamouda, D., Ferrag, M., Benhamida, N., & Seridi, H. (2022). PPSS: A privacy-preserving secure framework using blockchain-enabled federated deep learning for Industrial IoTs. *Pervasive Mob. Comput.*, 88, 101738. https://doi.org/10.1016/j.pmcj.2022.101738
19. Salim, S., Turnbull, B., & Moustafa, N. (2024). A Blockchain-Enabled Explainable Federated Learning for Securing Internet-of-Things-Based Social Media 3.0 Networks. *IEEE Transactions on Computational Social Systems*, 11, 4681-4697. https://doi.org/10.1109/tcss.2021.3134463
20. Khan, A., Laghari, A., Li, P., Dootio, M., & Karim, S. (2023). The collaborative role of blockchain, artificial intelligence, and industrial internet of things in digitalization of small and medium-size enterprises. *Scientific Reports*, 13. https://doi.org/10.1038/s41598-023-28707-9

Design and Performance Enhancement of a Multiband 1×4 Microstrip Array Antenna Using a Slotted Square-Ring Frequency-Selective Surface for Sub-6 GHz 5G Applications

D. M. Kumbhar[1], S.S. Ittannavar[2], S. S. Patil[3], D. B. Madihalli[4], S.J. Patil[5], Sanket Badakar[6]

[1.] *Research Scholar, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: dukumbhar.ece@hsit.ac.in.*

[2.] *Associate Professor, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: ssittannavar.ece@hsit.ac.in.*

[3.] *Assistant Professor, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: sachinpatil.ece@hsit.ac.in.*

[4.] *Assistant Professor, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: dbmadihalli.ece@hsit.ac.in*

[5.] *Assistant Professor, EEE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: sjpatil.ece@hsit.ac.in.*

[6.] *UG Scholar, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: sanketbadakar2004.ece@gmail.com.*

Abstract- In this work, a wideband 1×4 corporate-fed microstrip array antenna integrated with a slotted square-ring Frequency-Selective Surface (FSS) reflector for sub-6 GHz 5G application is presented. The design was optimized by performing parametric studies of the radiating elements, feed network, FSS reflector, and 1/4 wavelength air gap as shown in the ANSYS HFSS simulation. The antenna functions over the band of 3.5–5.8 GHz and has a high fractional bandwidth of 51.12%. It achieves a peak realized gain of 12.4 dBi with minimal gain variation (~1 dBi) over the band. The FSS reflector has high forward radiation and low back radiation and has the ability to increase the radiation efficiency to 77.5% without increasing the complexity of the feeds. This low cost and high performance design is applicable to 5G base stations, wireless backhaul and broadband directional communication systems.

Keywords - Frequency-Selective Surface (FSS); Microstrip Array Antenna; Radiation Enhancement; Sub-6 GHz.

1. INTRODUCTION

The explosive growth of wireless data, caused by the emergence of multimedia services, cloud computing, IoT, and M2M has fueled the development of 5G networks. Good propagation, wide coverage and non-line-of-sight transmission makes sub-6 GHz the preferred choice for 5G, and is ideal for macro-cellular base stations, wireless backhaul, industrial automation and broadband access networks. The antennas in this band thus need to have a wide bandwidth, high gain, stable radiation pattern, high efficiency, low cost and compact size [1]–[4]. Microstrip patch antennas are much used due to their low-profile, light weight, and ease of fabrication, but with a narrow bandwidth and moderate gain, single elements have limitations. Various enhancement techniques have been considered including slot loaded radiators, corporate fed arrays, DGS, EBG structures, AMC, and FSS, with corporate fed arrays and FSS reflectors showing good performance in terms of gain and radiation efficiency [5]–[12]. Difficulties exist however: For slot loading at least a moderate gain improvement is available; For DGS/EBG techniques the effect on the radiation performance is not very high; Most high gain FSS based designs require complex multi-layer structures and substrates that are expensive, and simultaneous wide bandwidth, stable gain, and high efficiency would be difficult to achieve [13]–[20]. To overcome these drawbacks, in this paper, a 1×4 (one by four) corporate-fed microstrip array antenna with slotted square-ring FSS reflector is designed for sub-6 GHz 5G applications using low-cost FR-4 substrate and simulated in ANSYS HFSS. The design provides a planar structure with simple design and cost savings, and enhances the impedance matching and radiation characteristics by using the corporate-fed array and FSS reflector. Based on the simulation results, the fractional bandwidth of the device is 51.12% (3.5-5.8 GHz), the peak realized gain of the device is 12.4 dBi and the radiation efficiency of the device is ~77.5%, making it suitable for broadband sub-6 GHz wireless applications [21]–[25].

2. LITERATURE REVIEW

The research of microstrip array antennas with a wide impedance bandwidth, high gain and stable radiation pattern has been stimulated by the requirement of high-rate wireless communication. To get high directivity and aperture efficiency, corporate-fed arrays are commonly used and to get better impedance matching, surface wave suppression, and radiation efficiency, various techniques were developed like DGS, EBG, AMC, and FSS

[7]–[12] are used. FSS reflectors are interesting for their frequency-selective properties, which enhance the forward radiation while reducing the back radiation, having a low profile and easy planar implementation when compared to conventional metallic reflectors. Most of the reported FSS-assisted array antennas, however, have a drawback of narrow bandwidth, gain fluctuations, multi-layer structure, or high cost of low-loss substrate [13]–[18] which make the fabrication more complicated. Recent studies on wideband microstrip array antennas [11], [13], [15], [17], [19] and [22]-[24] discuss the trade-offs between bandwidth, gain, structural complexity and profile ever since. Based on these results, this work investigates a corporate-fed 1×4 microstrip array with an optimized slotted square-ring FSS reflector for sub-6 GHz 5G applications. The design combines a joint optimization of the radiating elements, the feed network, the unit cells of the reflector, and the quarter-wavelength air gap, which results in better impedance bandwidth, high gain stability, high radiation efficiency and better directionality, while maintaining a low-cost, FR-4 implementation for practical wireless infrastructure.

3. ANTENNA DESIGN METHODOLOGY

3.1 Design Requirements and Antenna Configuration

The service of 5G wireless communication for sub-6 GHz band is also expected to become reality, which has also increased the demand for antenna systems to achieve a large impedance bandwidth, large realized gain, stable radiation characteristics, high radiation efficiency and a compact and low cost structure. The conventional microstrip patch antenna has low profile, lightweight construction and easy fabrication, but has poor bandwidth and medium gain, which limits the application of the broadband wireless system [21]-[22].

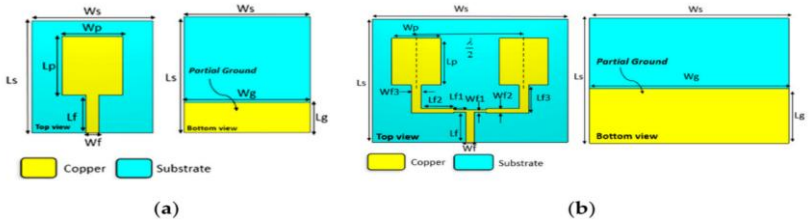


Figure 1. Geometry of the proposed 1×4 corporate-fed microstrip array antenna (a) single element, (b) 1 × 2 array antenna

The designs of the proposed antenna are shown in Fig. 1. The radiating array is made on a FR-4 substrate and fed in a corporate feed system for equal power distribution between the four radiating elements. An FSS reflector is located under the array with the air gap of about λ/4 to increase the forward radiation by constructive electromagnetic reflection. The antenna has been designed to cover the frequency range of 3.5 GHz to 5.8 GHz and analyzed using ANSYS HFSS. The proposed antenna in this design comprises of a 1x4 corporate-fed microstrip array that is integrated with a slotted square-ring Frequency-Selective Surface (FSS) reflector as shown in Fig. 2. The first configuration is a single rectangular microstrip patch with the desired resonant property, but has a small radiating aperture and thus low gain [22] and [23]. The antenna is designed as a 1x4 linear arrays for better directivity and the effective gain of the antenna is increased, thus improving the broad side radiation feature. These radiating elements are then fed into a corporate feed network to achieve equal A and B feeding, thus improving the match and power distribution. Then a slotted square-ring FSS reflector is added below the array at a frequency that is about λ/4 apart. The FSS enhances the forward radiation by positive reflection and attenuates the back radiation, which leads to better realized gain and front-to-back ratio and radiation efficiency with a lower level of antenna complexity [23].

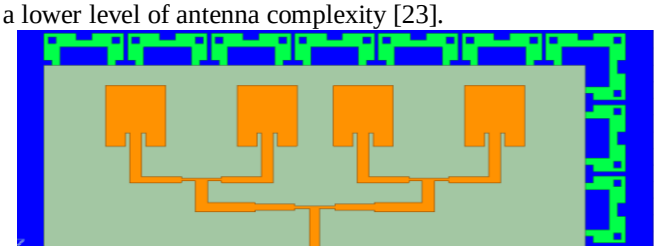


Figure 2. Design of the Proposed 1x4 Array Antenna with FSS Reflectors separated with an Airgap of 16.7 mm.

3.2. Frequency-Selective Surface Design and Performance Enhancement

The geometry of the proposed slotted square-ring FSS is illustrated in Fig. 3. The FSS is composed of a series of unit cells arranged in square rings with a specific periodicity, which gives it frequency selective reflecting properties. These changes to the effective current path allow the reflector to be designed for stable reflection in the operating band of sub-6 GHz [24-27].

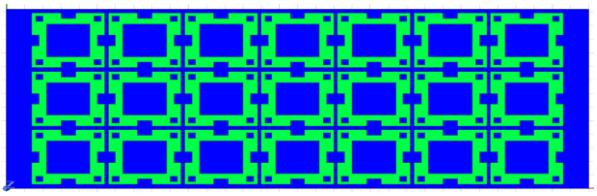


Figure 3. Design of the proposed FSS reflector with 3 × 7 SSR arrays.

This is done by placing the FSS reflector about λ/4 below the radiating array so as to create constructive interference between the reflected and radiated electromagnetic waves. This configuration increases forward gain, less back gain and better front-to-back ratio. The air-gap spacing, periodicity and dimensions of the FSS were varied to perform the parametric study in the simulation software, ANSYS HFSS, for achieving stable impedance matching and radiation characteristics within the operating band of 3.5-5.8 GHz [25].

3.5 Geometrical Parameters of the Proposed Antenna

The geometrical parameters of the proposed antenna are given in Table I. The antenna is a corporate fed microstrip array with FR-4 substrate under which a slotted square ring FSS reflector is located that reflects the incident wave. The antenna is designed with a 1 × 4 corporate fed microstrip array on FR-4 substrate and a slotted square ring FSS reflector is located below the radiating structure. To get constructive electromagnetic reflection, an approximately λ/4 air gap is kept between the antenna and the reflector.

Table I Geometrical Parameters of the Proposed Antenna.

Parameter	Wg	S	Lf1	Wf1	Wf2	Lf2	Wf3	Lf3
Value(mm)	136	16.88	5	0.79	1.78	10	2.745	15
Parameter	Ls	Ws	h	Lp	Wp	Lf	Wf	Lg
Value(mm)	55	136	1.6	18.25	15	11.84	2.745	29

4. Results and Discussion

4.1 Reflection Coefficient and VSWR

The reflection coefficient of the designed antenna (S_{11}) is shown in Fig. 4. The antenna has a reflection coefficient of less than −10 dB and an impedance bandwidth of 3.5–5.8 GHz, which is 51.12% fractional bandwidth. This better impedance matching is believed to be due to the combination of the corporate-fed array and the slotted square-ring FSS reflector to improve the electromagnetic coupling and minimize the reflection loss [21, 22]. The VSWR response and Measured reflection coefficient plot are shown in Fig. 5 and Fig. 6. The VSWR is maintained low over the band showing good match between the feed network and radiating elements with low mismatch losses. The results confirmed that the proposed antenna is suitable for sub-6 GHz wireless communication broadband application.

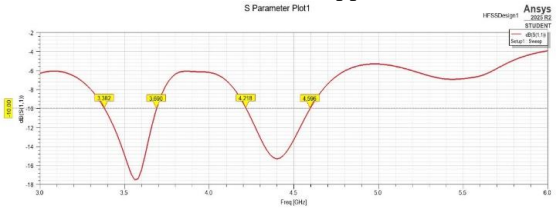


Figure 4. The simulated reflection coefficient plot of the 1x4 array antenna.

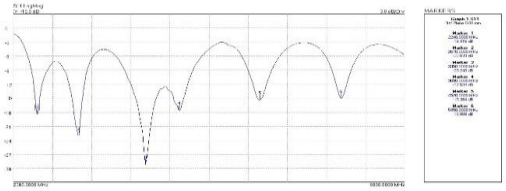


Figure 5. The Measured reflection coefficient plot of the 1x4 array antenna.

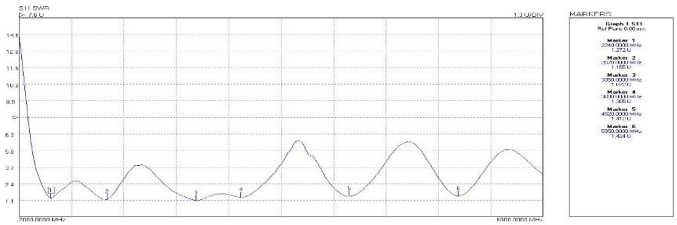


Figure 6. Measured VSWR Characteristics showing the current distribution well below the range.

4.2 Realized Gain and Radiation Efficiency

The simulated realized gain characteristics and fabricated antenna are shown in Fig. 7 and Fig. 8. The realized gain is 12.4 dBi with only slight variation over the operating band and thus shows good radiation stability. In addition to the electromagnetic reflection effect of the slotted square-ring FSS reflector [23], the gain enhancement is obtained from the effective aperture of the 1 × 4 corporate-fed arrays. The simulated radiation efficiency is about 77.5%, which is a good conversion efficiency of the input power to radiated power. The FR-4

substrate has dielectric losses, but the FSS reflector helps to increase the forward radiation and suppresses unwanted back radiation to compensate for these losses.

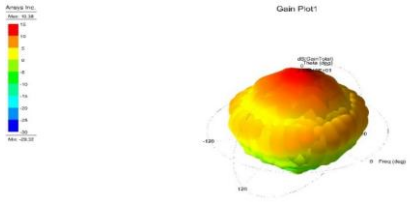


Figure 7. The Gain Analysis plot of proposed antenna.



Figure 8. The Photograph of Fabricated Antenna.

5. CONCLUSION

This work introduces a 1×4 corporate-fed microstrip array antenna that is integrated with a slotted square-ring FSS reflector for sub-6 GHz 5G applications. The antenna has been designed and analyzed using the software, ANSYS HFSS, and it has a wide operating bandwidth of 3.5–5.8 GHz (51.12% fractional bandwidth), a maximum realized gain of +12.4 dBi and a radiation efficiency of ~77.5%. The FSS reflector improves the forward radiation, reduces the back radiation and keeps the radiation characteristics stable over the band. The proposed antenna features a broad bandwidth, high gain, and efficient radiation efficiency on a low cost FR-4 substrate, which is suitable for sub-6 GHz 5G base stations, wireless backhaul and broadband wireless systems. The prototype fabrication, experimental validation and usage of different loss of substrates and advanced FSS geometry will be undertaken in future studies to further improve the performance.

6. REFERENCES

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*, 4th ed. Hoboken, NJ, USA: Wiley, 2016.

[2] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook*. Norwood, MA, USA: Artech House, 2001.

[3] D. M. Pozar, "Microstrip antennas," *Proc. IEEE*, vol. 80, no. 1, pp. 79–91, Jan. 1992.

[4] D. M. Pozar and D. H. Schaubert, *Microstrip Antennas: The Analysis and Design of Microstrip Antennas and Arrays*. New York, NY, USA: IEEE Press, 1995.

[5] F. Yang and Y. Rahmat-Samii, *Electromagnetic Band Gap Structures in Antenna Engineering*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[6] B. A. Munk, *Frequency Selective Surfaces: Theory and Design*. New York, NY, USA: Wiley, 2000.

[7] N. Kushwaha and R. Kumar, "Design of slotted microstrip patch antenna for broadband applications: A review," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 29, no. 1, 2019.

[8] S. Koziel and A. Bekasiewicz, *Antenna Design by Simulation-Driven Optimization*. Cham, Switzerland: Springer, 2019.

[9] H. Wong, K.-M. Luk, C. H. Chan, Q. Xue, K. K. So, and H. W. Lai, "Small antennas in wireless communications," *Proc. IEEE*, vol. 100, no. 7, pp. 2109–2121, Jul. 2012.

[10] D. Sievenpiper, "High-impedance electromagnetic surfaces," Ph.D. dissertation, Univ. California, Los Angeles, CA, USA, 1999.

[11] A. A. Kishk, "Wideband and high-gain antennas for wireless communications," *IEEE Antennas Propag. Mag.*, vol. 60, no. 6, pp. 8–20, Dec. 2018.

[12] H. Oraizi and S. Hedayati, "Miniaturization of microstrip antennas by the novel application of the Giuseppe Peano fractal geometries," *IEEE Trans. Antennas Propag.*, vol. 60, no. 8, pp. 3559–3567, Aug. 2012.

[13] Y. Li, Z. N. Chen, X. Qing, Z. Zhang, J. Xu, and Z. Feng, "Axial-ratio bandwidth enhancement of patch antennas using parasitic elements," *IEEE Trans. Antennas Propag.*, vol. 57, no. 9, pp. 2783–2787, Sep. 2009.

[14] K.-L. Wong, *Compact and Broadband Microstrip Antennas*. New York, NY, USA: Wiley, 2004.

[15] J. Huang and J. A. Encinar, *Reflectarray Antennas*. Hoboken, NJ, USA: Wiley-IEEE Press, 2008.

[16] F. Costa, S. Genovesi, and A. Monorchio, "A frequency selective radome with wideband absorbing properties," *IEEE Trans. Antennas Propag.*, vol. 60, no. 6, pp. 2740–2747, Jun. 2012.

[17] R. Gonzalo, P. de Maagt, and M. Sorolla, "Enhanced patch antenna performance by using electromagnetic band-gap substrates," *IEEE Microw. Wireless Compon. Lett.*, vol. 9, no. 10, pp. 403–405, Oct. 1999.

[18] D. Sievenpiper, L. Zhang, R. Broas, N. Alexopolous, and E. Yablonovitch, "High-impedance electromagnetic surfaces with a forbidden frequency band," *IEEE Trans. Microw. Theory Techn.*, vol. 47, no. 11, pp. 2059–2074, Nov. 1999.

[19] F. Yang and Y. Rahmat-Samii, "Reflection phase characterizations of the EBG ground plane for low profile wire antenna applications," *IEEE Trans. Antennas Propag.*, vol. 51, no. 10, pp. 2691–2703, Oct. 2003.

[20] S. Maci and G. B. Gentili, "Dual-frequency patch antennas," *IEEE Antennas Propag. Mag.*, vol. 39, no. 6, pp. 13–20, Dec. 1997.

[21] A. Vallecchi, F. Capolino, and F. De Flaviis, "Low profile fully planar folded dipole antenna on a high impedance surface," *IEEE Trans. Antennas Propag.*, vol. 60, no. 1, pp. 51–62, Jan. 2012.

[22] T. Debogović, J. Perruisseau-Carrier, and J. R. Mosig, "Wideband printed reflectarray antennas," *IEEE Trans. Antennas Propag.*, vol. 60, no. 10, pp. 4520–4526, Oct. 2012.

[23] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 183–198, Jan. 2014.

[24] S. Costanzo, F. Venneri, G. Di Massa, and A. Borgia, "Hybrid reflectarray antenna for broadband wireless applications," *IET Microw., Antennas Propag.*, vol. 7, no. 7, pp. 526–532, 2013.

[25] S. K. Sharma and L. Shafai, "Performance enhancement of microstrip patch antennas using frequency selective surfaces: A review," *IEEE Access*, 2021.

[26]. A. B. Arabhanvi, S. S. Ittannavar, J. S. Jakati, V. Patil, S. Ranbhare and B. A. P. Khot, "High-Performance UWB Microstrip Antenna Using Gap-Coupled Suspended Technique for 5G Connectivity," 2026 International Conference on Computing Theory and Wireless Communications (ICCTWC), Ichalkaranji, India, 2026, pp. 1-6, doi: 10.1109/ICCTWC68241.2026.11583436.

[27] D. M. Kumbhar, S. S. Ittannavar, R. S. Hosamani, R. Hosamani, B. A. Pratima Khot and A. B. Arabhanvi, "Dual Band Microstrip Patch Antenna Using Compact Structural Tuning for Smart Wireless Applications," 2025 2nd Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2025, pp. 1-5, doi: 10.1109/APCIT65661.2025.11410640.

<IC-DAIS>
<Conference Series name> <Volume number and Year> <DOI Number>

Factors Affecting Customers Satisfaction in the Airline Industry using AHP and TOPSIS Methods

Trupthi Rao

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India trupthirao-aiml@dsu.edu.in

Souramita Bhowmik

Dept. of CSE (DS), Dayananda Sagar University, Bangalore, Karnataka, India souramita-ds@dsu.edu.in

Shivnandan Rai

Dept. of CSE (AI&ML), Dayananda Sagar University, Bangalore, Karnataka, India shivnandanrai-aiml@dsu.edu.in

Ashwini Kodipalli

School of CSE (AI&ML), PES University, Bangalore, Karnataka, India ashwinik@pes.edu

Shridhar B Devamane

Dept. of CSE, MS Engineering College, Bangalore, Karnataka, India devamaneshridhar@gmail.com

Gargi N

Dept. of AI & Data Science, Global Academy of Technology, Bangalore, Karnataka, India gargi@gat.ac.in

Abstract.

In the airline industry customer satisfaction varies with respect to multiple features such as price, quality, safety, attentivity, punctuality. In this research we propose an integrated AHP, TOPSIS modelling to find the best features that contribute to decision making. Also, these features help in evaluating airline customer satisfaction. AHP method finds the feature importance. TOPSIS assigns rank to the alternatives by considering the distance of alternative from Positive and Negative Ideal Solutions. Thereby, finding the Closeness Coefficient. Experimental results show that service quality and flight experience are of utmost importance in the customer satisfaction decision. The framework offers airlines a data-driven tool for prioritizing service improvements.

Keywords. Customer satisfaction, Airline industry, Fuzzy TOPSIS, Fuzzy AHP, Fuzzy Synthetic Evaluation, Closeness Coefficient

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

In today's competitive airline industry, customer satisfaction is crucial for success, as rising expectations and low-cost carriers increase competition. Customer satisfaction is based on several factors that includes safety, food/beverage availability, punctuality, cabin comfort, services provided. Satisfaction depends on factors such as punctuality, cabin comfort, entertainment, food quality, and customer service. Traditional methods fail to capture uncertainty of data because customer preferences can keep changing. Fuzzy methods can deal with uncertain data and hence become an effective approach in the selection of features that help in decision making of a model. With the help of these methods airline industry officials can provide better facility to passengers and gain customer confidence, satisfaction. This paper includes literature review, proposed methodology, results and implications.

2. RELATED WORK

Previous studies have identified service quality, price, flight attendee behaviour as primary features for airline customer satisfaction (Wang'ondu, 2009 [8]). Research carried out by Saha et al. (2016) [9], Wilson & Christella (2019) [10], and Rahman (2014) [11] included more features such as responsiveness, assurance, and reliability. Singh (2016) [7] showed the use of AHP in concluding air industry service quality features, while Garg (2020) [3] and Mahtani & Garg (2018) [2] used FAHP to handle the uncertainty in airport/airline decision-making. Raut et al. (2016) [1] used CFA, FAHP, TOPSIS for airline service-quality evaluation and determined cabin, booking and check-in service as important attributes. Navid (2017) [6] proposed Type-2 Fuzzy TOPSIS to incorporate ambiguity in passenger evaluations and successfully ranked Iranian airlines. Saeida Ardakani et al. (2015) [4] used fuzzy methods to identify service-quality gaps, while Hassanzadeh & Asghari (2020) [5] applied fuzzy AHP-TOPSIS for operational-factor ranking. Kumar et al. (2024) [12] demonstrated the applicability of fuzzy AHP-TOPSIS to cybersecurity vulnerability assessment, highlighting its broader decision-making potential. Kumar et al. (2024) [13] incorporated machine learning for airline customer-satisfaction prediction, indicating a shift toward data-driven approaches.

3. METHODOLOGY

This project employs basic Fuzzy methods such as fuzzy AHP (FAHP), FAHP using a geometric means and TOPSIS methods for selecting the features for customer satiafaction. We define eight essential attributes: price (c1), Range kmph (c2), Capacity (c3), Seats (c4), Variants (c5), comfort (c6), and fuel type(c7) The outcome is a prioritized ranking of all these allocation options, providing valuable guidance for selecting a car brand. This study is to determine the features/factors that affect the customer satisfaction in airline. These systems influence various algorithms to identify the most critical factors affecting customer satisfaction, airline operators can focus on improving these areas to enhance overall customer experience.

3.1. Analytic Hierarchy Process (AHP)

Thomas L. Saaty developed AHP in the 1980s. It is an organized method of overseeing and evaluating complex choices that comprise segmenting a decision problem into more manageable subproblems that can be independently examined. The hierarchy elements can

relate to any condition of the decision problem, such as definite or indefinite conditions, or quantitative or qualitative factors. This method is generally used in various fields including business, healthcare, and engineering.

3.2. TOPSIS

The second method that is employed here is the TOPSIS method. In this method, each criterion is rated on a 5-point linguistic scale: Service Quality ranges from Excellent (5) to Very Poor (1); Flight Experience from Outstanding (5) to Unsatisfactory (1); Pricing from Very Affordable (5) to Very Expensive (1); and On-time Performance from Always On Time (5) to Always Late (1).

4. RESULTS

4.1. AHP Method

The attribute with the highest criteria weight, 0.3587 (Price), is identified as the primary selecting factor, while the attribute with the lowest criteria weight, 0.0393 (Fuel Type), is determined to be the least influential factor in the selection process. Table I shows the final results after performing all calculation.

TABLE I. FINAL RANKING OF ALTERNATIVES

Alternatives	Service Quality	Flight Experience	Price	On-time Performance
A1	5	4	3	5
A2	4	5	4	4
A3	3	3	2	3

The Alternative (A) with the highest CC*i* value is provided the highest rank. Table II shows the closeness coefficient of the alternatives.

TABLE II. CLOSENESS COEFFICIENT OF ALTERNATIVES

Alternatives	Closeness Coefficient (CC*i*)
A1	0.666
A2	0.451
A3	0

The Alternative with the highest CCi score and with optimal solution was A1, with a C of 0.666 which has best ratings for customers satisfaction and highly recommended followed by A2 which as a moderate rating and then A3 with the least rating is not all recommended.

5. CONCLUSION

By using AHP method we produced a hierarchy of important features/criteria that included the comfort of seats, on-boarding service, booking in online mode, availability of food, drinks and handling of bags. With the help of pairwise matrix, we got to identify the most crucial features. Also, the consistency index helped in identifying the important features with their associated weights. With the help of TOPSIS method, rank was given to various criteria. The features with the highest closeness coefficient got the highest rank. Ranking was done with respect to their distance from the positive ideal and negative ideal solutions. Comfortable seating and high-quality customer service were the important features when combined both the methods.

6. REFERENCES

[1] Raut, R. D., Kamble, S. S., & Jha, M. K. (2016). A combined confirmatory factor analysis and fuzzy-analytic hierarchy process-TOPSIS (CFAT) framework for measurement of airline service quality. International Journal of Services and Operations Management, 23(3), 347-386.

[2] Mahtani, U. S., & Garg, C. P. (2018). An analysis of key factors of financial distress in airline companies in India using fuzzy AHP framework. Transportation Research Part A: Policy and Practice, 117.

[3] Garg, C. P. (2020). Evaluation of the airport service quality indicators using fuzzy AHP framework: An Indian perspective. International Journal of Services and Operations Management, 37(3), 320-342.

[4] Saeida Ardakani, S., Nejatian, M., Farhangnejad, M. A., & Nejati, M. (2015). A fuzzy approach to service quality diagnosis. Marketing Intelligence & Planning, 33(1), 103-119.

[5] Hassanzadeh, R., & Asghari, H. (2020). Identification and ranking of affecting factors on sales and operations planning (S&OP) process implementation by using fuzzy AHP and fuzzy TOPSIS approach (case study: dairy industry). Journal of applied research on industrial engineering, 7(1), 57-78.

[6] Navid, H. (2017). Proposing a framework for airline service quality evaluation using Type-2 Fuzzy TOPSIS and non-parametric analysis. Journal of Sustainable Development of Transport and Logistics, 2(2 (3)), 6-25.

[7] Singh, A. K. (2016). Competitive service quality benchmarking in airline industry using AHP. Benchmarking: An International Journal, 23(4), 768-791.

[8] Wang'ondu, R. W. (2009). Factors affecting customer satisfaction in airline industry (Doctoral dissertation).

[9] Saha, N., Islam, M. R., & Hoque, A. U. (2016). Factors affecting customers' satisfaction of mobile phone subscribers: an empirical study on mobile telecommunication Industry in Bangladesh. International Journal of Business and Management, 11(6), 252.

[10] Wilson, N., & Christella, R. (2019). Empirical research of factors affecting customer satisfaction: A case of the Indonesian e-commerce industry. DeReMa Jurnal Manajemen, 14(1), 21-44.

[11] Rahman, H. (2014). Factors affecting customer satisfaction in mobile telecommunication industry in Bangladesh. Business, Management and Education, 12(1), 74-93.

[12] Kumar, A., Kodipall, A., Rao, T., Soma, B., & Gargi, N. (2024, April). Enhancing Password Security: Fuzzy AHP and TOPSIS-based Vulnerability Assessment. In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

[13] Kumar, A., Mehta, V., Rao, T., & Kodipalli, A. (2024, April). Forecasting Airline Customer Satisfaction through the Application of Diverse Machine Learning Algorithms. In *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)* (pp. 1-5). IEEE.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Intelligent Domain-Specific Content Categorization Framework Using a Deep Learning-Based Hybrid Approach

V.K. Janani Sivapriya[1,2], R. Saravana Moorthy[2], S. Angel[3]
[1]Department of Data Science, Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai, India
[2]Department of Computer Science, Kongunadu Arts and Science College, Coimbatore, India
[3]Department of Cybersecurity with CS, PSGR Krishnammal College for Women, Coimbatore, India

Abstract.

Automatic content classification is becoming increasingly important as the volume of multi-domain digital information produced by business, healthcare, education, sports, and technology grows exponentially. Traditional text categorisation methods are unable to capture semantic relations within domain-specific documents, resulting in low classification performance when applied across domains. To solve this problem, this paper proposes an intelligent multi-domain content categorization system. First, Multi-Level Contextual Text Refinement Filtering (MCTRF) performs denoising, text normalisation, domain-term preservation, and context-consistency enhancement to obtain a high-quality text representation through a comprehensive text refinement process. Thereafter, a Graph Attention Network (GAT) is used to capture and learn the semantic correlations among domain-specific concepts by assigning different attention weights to neighbouring nodes, thereby yielding informative domain-relationship features. Lastly, a hybrid Convolutional Neural Network with a Bidirectional Long Short-Term Memory (CNN-BiLSTM) classifier is applied to capture both local semantic patterns and long-term contextual information for precise content classification. The evaluation is performed on a public multi-domain text dataset using the parameters are accuracy, precision, recall, and F1-score. It is anticipated that the presented framework will provide better results than traditional approaches, and deliver a computationally efficient, scalable approach for multi-domain document categorisation.

Keywords. Deep Learning (DL), Domain based content categorisation, relationship, hybrid approach, local features, global features.

1. INTRODUCTION

The boom in digital text across business, healthcare, education, sports, entertainment, and technology presents the challenge of developing smart, domain-specific content classification systems. Automatic domain label extraction is an important task for information retrieval, recommender systems, knowledge management, and decision support, facilitating the organisation and analysis of large-scale textual data (Yuyuan Zhang et al., 2025). Conventionally, domain classification is based on expert manual annotation, a process in which a set of predefined labels is assigned to a document based on its semantics. However, such a process is time-consuming, labour-intensive, costly and inconsistent, and

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

thus it is not feasible for large-scale and real-time applications. Therefore, automatic domain classification is an efficient way to handle diverse textual data. Recent progress in DL has led to dramatic improvements in the performance of automatic text classification by learning hierarchical semantic representations from raw document input. DL-based document classification methods also effectively capture sequential context and have been shown to outperform traditional machine learning methods for document classification in various domains such as Technology, Business, Sports, Entertainment, and Politics D. Tarun et al., (2023).

However, current DL techniques are prone to producing redundant document representations, failing to exploit the semantic relations within domain knowledge fully, and struggling to model both local and global contextual patterns. These constraints degrade the ability to classify domain-specific data accurately. To overcome these challenges, we introduce an intelligent domain-specific content categorisation system based on a hybrid DL model in this paper. The framework utilises MCTRF for contextual denoising, lexical normalisation, and the retention of domain-specific keywords, and then employs a GAT model to capture semantic relations among domain terms. Finally, a Hybrid CNN–BiLSTM classifier is proposed to combine local feature extraction with bidirectional contextual learning to achieve a more accurate and robust multi-domain content classifier. The primary contributions of this study are as follows:

- A novel MCTRF framework is proposed for textual denoising, lexical normalisation, contextual refinement and domain-specific keyword preservation.
- To produce relation-aware feature representations, a GAT-based semantic learning model is proposed that adaptively attends to different parts of the input data.
- A Hybrid CNN–BiLSTM classifier is proposed to simultaneously capture local semantic features and long-range contextual dependency for effective multi-domain text classification.
- The proposed framework is tested on a public multi-domain text dataset and compared in terms of accuracy, precision, recall, and F1 Score.

2. LITERATURE SURVEY

The research in L. Liu *et al*., (2024) proposed an Improved Text Label Optimization (ITLO) framework for Specific Domain Prompt Learning (SDPL) to optimize prompt representations for domain-specific text classification. The research in D. K. Meghana *et al*., (2025) introduced an Auto-Encoder-Based Structural Correspondence Learning with Pivot Augmentation (AESCL-PA) model for multilingual text classification. The work in Emanuel, R. H *et al*., (2024) evaluated different activation functions in Convolutional Neural Network (CNN) models for text classification. The approach in Sun, Z *et al*., (2024) developed a Digital Twin with Unsupervised Domain Adaptive Learning (DT-UDAL) framework for fault diagnosis using Proton Exchange Membrane Fuel Cell (PEMFC) systems. The methodology in Zhu, J *et al*., (2024) proposed Causal Prototype-Inspired Contrast Adaptation (CPCA) for Unsupervised Domain Adaptive Semantic Segmentation (UDASS). The model reduced domain discrepancies using contrastive adaptation and achieved superior segmentation accuracy and cross-domain generalization. The study in Nair, A. R *et al*., (2024) investigated text data augmentation techniques using DistilBidirectional Encoder Representations from Transformers (DistilBERT) for text classification.

The framework in Yin, K *et al*., (2026) introduced CrisisSense-Large Language Model (CrisisSense-LLM) for multi-label disaster-related social media text classification. The instruction fine-tuned model effectively captured contextual semantics and achieved high precision, recall, and F1-score. The research in Sy, C. Y *et al*., (2024) compared Robustly Optimized Bidirectional Encoder Representations from Transformers Approach (RoBERTa) for supervised text classification. The method in Zheng, T *et al*., (2024) proposed Character Distance Network (CDistNet) for robust multi-domain text recognition. The framework learned character-distance relationships across domains and improved recognition accuracy under diverse conditions. The study in I. Bagwan *et al*., (2026) presented Domain-Adaptive Transfer Learning (DATL) for scam call detection. The proposed model enhanced transferability across domains while maintaining privacy and achieved higher detection accuracy.

3. PROPOSED METHODOLOGY

Our proposed framework operates on a domain-based dataset of Business, Technology, Sports, Entertainment, and Education, as presented in figure 1. The MCTRF module first preprocesses the documents by dropping URLs, punctuation, and other noise, and then performs stop word removal and lemmatisation to obtain a polished textual representation.

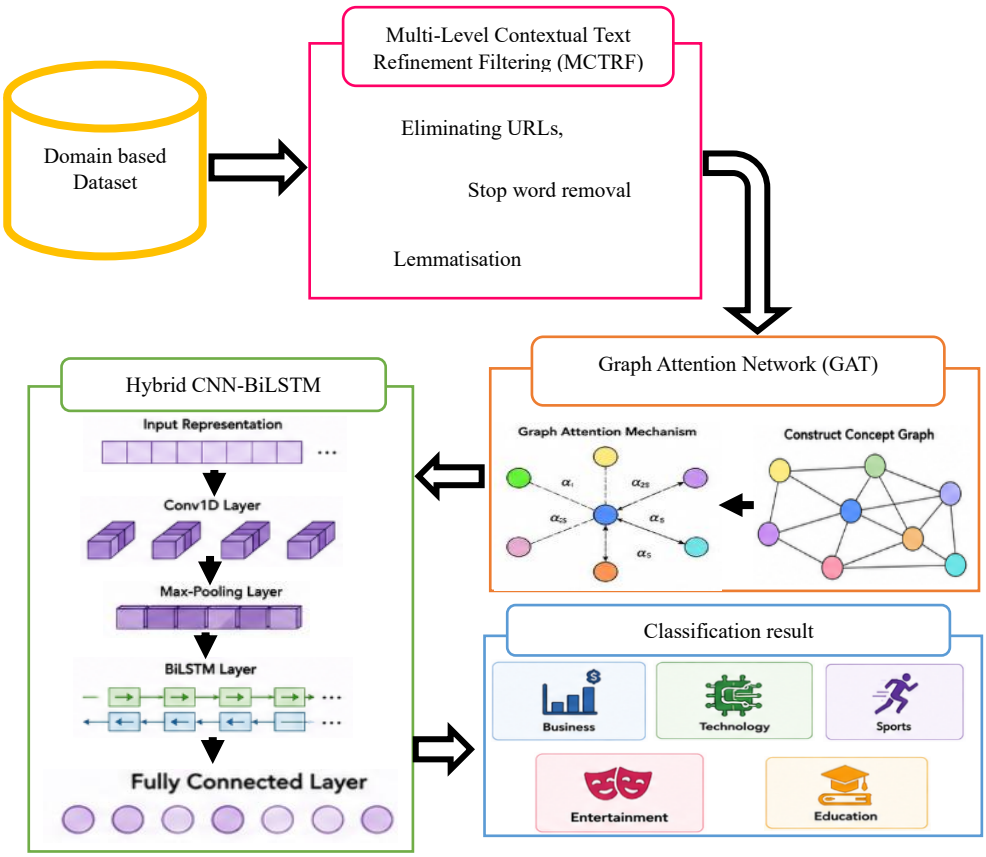


Figure .1. Architecture diagram

The refined documents are then passed to the GAT to establish semantic relations between domain-related keywords by assigning attention weights to neighbouring concepts. As a result, they produce a relation-aware feature representation that encodes both local and contextual semantic information. The relationship-aware features, in turn, are fed into the Hybrid CNN–BiLSTM classifier, where the CNN learns local semantic features and the BiLSTM learns bidirectional contextual relations. The final SoftMax layer predicts the domain category.

3.1 Multi-Level Contextual Text Refinement Filtering (MCTRF)

The presented MCTRF approach enhances the semantic quality of the multi-domain news dataset, including articles on Business, Technology, Sports, Education, and Entertainment. Each news item has a headline, a description, and the article content, which are initially combined into a single document to preserve the full context. The integrated document is represented as equation 3.1.

$$T_i = h_i \oplus s_i \oplus c_i \tag{3.1}$$

Here, T_i refers to integrated textual document, h_i lies the headline, s_i express the description, c_i denotes the article content, and $\oplus$ denotes Concatenation operator. The merged document is then preprocessed by eliminating URLs, punctuation, special characters, numbers, and other irrelevant content to denoise the data and enhance semantic uniformity. The refinement process is given by equation 3.2.

$$T_i^r = \phi(T_i) \tag{3.2}$$

Where, T_i^r refined document and $\phi(\cdot)$ is the Noise removal function. Then the processed text is tokenised into a sequence of words, and the words are lowercased to reduce lexical discrepancies across documents.

$$W_i = \{w_1, w_2, \dots, w_m\} \tag{3.3}$$

Herein, W_i is the token sequence, w_j individual token, and m refers to number of tokens. The normalised tokens are then tokenised on whitespace to obtain a semantically meaningful sequence of tokens, after eliminating domain-irrelevant stop words.

$$W_i^s = W_i - S \tag{3.4}$$

Let us assume, W_i^s is the filtered token sequence, S is the stop word vocabulary. Following lemmatisation, several inflected forms are mapped to their canonical root forms, which helps reduce the effect of vocabulary sparsity while preserving the implicit meaning.

$$L_i = \psi(W_i^s) \tag{3.5}$$

Where, L_i is the Lemmatized token sequence, and $\psi(\cdot)$ lies the Lemmatization function. Lastly, keep domain-related keywords to strengthen and preserve the semantic representation for better domain discrimination before performing graph-based relationship learning.

$$R_i = L_i \cup K \tag{3.6}$$

whereR_i denotes the refined document, $K$is the domain specific keyword set, and $\cup$ is the union operator. The resultant refined document R_i, which contains semantically enriched domain-specific information after redundancy removal, is then sent to the GAT for learning semantic relationships.

3.2 Graph Attention Network (GAT)

After preprocessing, our proposed GAT captures the semantic correlations among domain-specific concepts by mapping each processed document to a semantic graph, in which keywords serve as nodes, and their contextual co-occurrences define edges. Each

node is initialised with a contextual feature vector and projected into a learnable latent space. The graph attention network now computes adaptive attention coefficients to weight information from adjacent nodes based on semantic relevance. Finally, the new node embeddings are integrated to form a relation-aware feature representation that captures not only local semantic interactions but also global contextual dependencies.

$$G = (V, E) \tag{3.7}$$

Here, G is the semantic graph, $V = \{v_1, v_2, \dots, v_n\}$ is the set of graph nodes representing domain specific keywords, $E \subseteq V \times V$ is the set of semantic edges representing contextual relationships. Above equation transforms the refined document into a semantic graph, where each node is a domain-specific keyword and edges represent contextual relationships among keywords. Using its contextual feature vector, each graph node is initialized

$$H = \{h_1, h_2, \dots, h_n\} \tag{3.8}$$

Herein, H denotes the node feature matrix, h_i is the Contextual feature vector of node v_i, and n is the feature dimension. Equation 3.8 computes each graph node is initialized with its contextual feature vector, which is the input to graph attention learning. Before attention computation, each node feature is projected into a learnable latent feature space.

$$z_i = W h_i \tag{3.9}$$

Where, z_i is the transformed node feature, W is the trainable weight matrix, and h_i refers to original node feature. Each node feature is transformed into a higher-dimensional latent representation using Equation 3.9, thereby helping the network learn more discriminative semantic information for computing attention weights. The semantic importance of neighbouring nodes is calculated as

$$e_{ij} = \mathrm{ReLU}\left(a^T [z_i \parallel z_j]\right) \tag{3.10}$$

where, e_{ij} is the unnormalized attention coefficient, a is the learnable attention vector, $[z_i \parallel z_j]$ is the concatenated neighboring node features, and ReLU is the activation function. The semantic relevance of two connected nodes is computed by equation 3.10. Its attention scores for two neighbours who have a stronger contextual relation are greater than those for neigbours with a weaker tone. The attention coefficients are normalised through the SoftMax function;

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \tag{3.11}$$

Equation 3.11 allows us to normalise the attention scores between all semantically connected nodes, so that their sum is 1, and it also enables us to emphasise the furthest relevant semantic neighbours in the network adaptively. Here, α_{ij} refers to normalized attention coefficient, and $N(i)$ is the neighbour set of nodes i. The normalised attention coefficients are used to obtain the new node embedding by aggregating neighbourhood information.

$$h_i' = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} W h_j\right) \tag{3.12}$$

Where, h_i' is the Updated node representation, $\sigma(\cdot)$ activation function, α_{ij} is the normalized attention coefficient, Wh_j denotes the transformed neighbouring feature. Semantic information is then accumulated from neighbouring nodes according to the attention weights learned by Equation 3.12. A node can retain its own semantic features and acquire contextual information from linked domain concepts. The updated node embeddings are then combined to form the final relation-aware feature matrix.

$$F_R = \{h_1', h_2', \dots, h_n'\} \tag{3.13}$$

Where, F_R is the relationship aware feature matrix and h_i' is the updated embedding of node i. The updated node embeddings are concatenated and multiplied by the matrix to

produce the final relation-aware feature representation. This feature matrix contains both local semantic interactions and global contextual dependencies among domain-specific concepts. It will be fed to a Hybrid CNN–BiLSTM classifier for high-precision multi-domain content classification.

3.3 Hybrid CNN-BiLSTM approach

Based on the feature representation, the proposed model employs a Hybrid CNN–BiLSTM architecture for domain classification. The relationship-aware feature matrix F_R is fed into the classifier, where the CNN learns local semantic patterns with 1D convolution, followed by ReLU and max-pooling, as shown in Fig.1. The pooled features are then fed into a BiLSTM to model the long-range contextual correlations. In the end, the contextualised representation passes through a fully connected layer, and the Softmax classifier predicts the probabilities of the domain classes; the one with the highest probability is selected as the output label. The output of GAT, the relationship-aware feature matrix, is denoted as

$$X = F_R = \{x_1, x_2, \dots, x_m\} \in \mathbb{R}^{m \times d} \quad (3.14)$$

Where, X is the feature matrix, x_i is the feature vector of the i^{th}relationship-aware feature, m denotes the number of selected relationship aware features, and F_R is the Relationship-aware feature representation. The relationship-aware feature matrix obtained from the GAT is represented as in Equation 3.14. These compacted semantic features are then fed into the Hybrid CNN–BiLSTM classifier.

$$C_i = \sigma\left(\sum_{j=0}^{k-1} W_j\, x_{i+j} + b\right) \quad (3.15)$$

where, $C_i$is the convolution feature map, W_j is the convolutional kernel, x_{i+j} is the input feature vector, k is the kernel size, b is the bias, and $\sigma(\cdot)$ activation function. Based on the neighboring relationship-aware vectors, local semantic features are extracted by the convolution layer. The 2D convolution of 2D input blocks with 2D filters and the 1D convolution of 1D input sequences with 1D filters are popular in many fields. The convolution feature maps are sub-sampled via Max-Pooling.

$$P_i = \max_{j \in \Omega_i} (C_j) \quad (3.16)$$

Here, P_i is the pooled feature, and Ω_i is the pooling window. The max activation within each pooling window is taken in Equation 3.16, which maintains the prominent semantic information while decreasing the feature size.The Sequence of pooled features is cursorily processed in the forward direction.

$$\overrightarrow{h_t} = LSTM\left(x_t, \vec{h}_{t-1}\right) \quad (3.17)$$

Equation 3.17 captures contextual dependencies by running a sequence from the left to the right. The reverse is true for the same sequence. Herein, x_t is the input feature at time step t, $\overrightarrow{h_t}$ is the forward hidden state, and $\vec{h}_{t-1}$ is the previous hidden state. The reverse direction also learns contextual information in Equation 18, where future semantic information can guide feature learning.

$$\overleftarrow{h_t} = LSTM\left(x_t, \overleftarrow{h}_{t+1}\right) \quad (3.18$$

Herein, $\overleftarrow{h_t}$ is the backward hidden state, and $\overleftarrow{h}_{t+1}$ is the next hidden state. The hidden states of the two directions are concatenated.

$$H_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \quad (3.19)$$

The forward and backward context representations are concatenated to form rich semantic features in 19. Herein, H_t is the Bidirectional contextual representation, and [;]is the Concatenation operator. The contextual representation is then turned into a discriminative feature vector.

$$Z = W_d H + b_d \tag{3.20}$$

Equation 3.20 is used to map the contextual features to a higher-level representation for multi-class classification. Here, Z is the dense layer output, W_d is the weight matrix, H lies the BiLSTM output, and b_d is the bias vector. The probability of every domain class is calculated as

$$P(y = c \mid Z) = \frac{\exp(W_c Z + b_c)}{\sum_{i=1}^{C} \exp(W_i Z + b_i)} \tag{3.21}$$

Where, $P(y = c \mid Z)$ is the probability of every domain class, W_c is the weight matrix of class c, b_c is the bias, and c is the total number domain class. The target class for prediction is

$$\hat{y} = \arg \max_{c} P(y = c \mid Z) \tag{3.22}$$

The posterior probability for each content-category is calculated by Equation (3.21), and then d is classified into the category with the maximum probability score by Equation (3.22). The ultimate output of the proposed framework is the predicted domain label along with its classification probability, where each news article is classified into one among five classes i.e. Business, Technology, Sports, Education, and Entertainment.

4. RESULT AND DISCUSSION

The framework was developed in Python using Visual Studio Code (VS Code). The experiments are performed on a public Kaggle multi-domain news dataset with 5 categories: Business, Technology, Sports, Entertainment, and Education. The dataset is split 70% for training and 30% for testing, ensuring consistent model evaluation. The presented solution Hybrid CNN–BiLSTM classifier is evaluated using accuracy, precision, recall, and F1-score. Two state-of-the-art methods, i.e., DATL and DT-UDAL, are used to evaluate the proposed method's performance and demonstrate its excellent domain-dependent content classification.

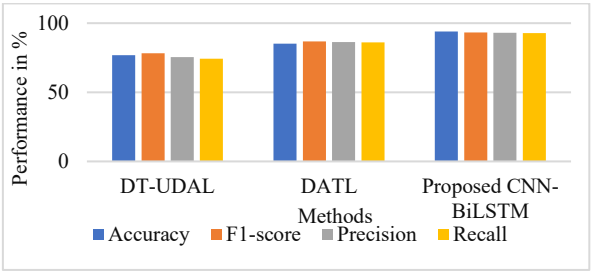


Figure.2. Comparison of accuracy and recall performance

Figure 2 presents a comparison of the accuracy, precision, recall, and F1-score between our CNN–BiLSTM model and DT-UDAL and DATL. DT-UDAL achieved 76.90%, 75.40%, 74.21%, and 78.31% in accuracy, precision, recall, and F1-score, respectively, and DATL further improved these results to 85.24%, 86.39%, 86.09%, and 86.90%. Our proposed CNN–BiLSTM achieved even better results, with an accuracy of 94.03%, a precision of 93.06%, a recall of 92.80%, and an F1-score of 93.27%, beating both methods.

5. CONCLUSION

In this paper, an intelligent domain-specific content categorisation using a DL-based Hybrid Approach is developed to classify multi-domain news articles into the Business, Technology, Sports, Entertainment and Education categories. In general, the proposed framework comprises three parts: The MCTRF module for contextual text pre-processing, GAT to capture semantic relations among domain-specific concepts, and a hybrid CNN–BiLSTM classifier to extract local semantic features and long-range contextual information. Results on a public Kaggle dataset demonstrate that the proposed framework attained the best performance, i.e., 94.03% for accuracy, 93.06% for precision, 92.8% for recall, and 93.27% for the F1 score, which significantly exceeded the state-of-the-art DT-UDAL and DATL approaches.

6. REFERENCE

[1] Yuyuan Zhang, Xiaolan Wang, Feiyan Liu, Yibin Jin, Zechen Liu, Kun Shi, Using domain-specific eyword features to enhance deep learning-based pressure vessel inspection problem identification, Engineering Applications of Artificial Intelligence, Volume 160, Part C, 2025.

[2] D. Tarun et al. “Domain Classification of Text Conversation using Deep Learning”, International Research Journal, Volume 10, Issue 7, pp.368-373, 2023.

[3] L. Liu, N. Wang, D. Liu, X. Yang, X. Gao and T. Liu, "Towards Specific Domain Prompt Learning via Improved Text Label Optimization," in *IEEE Transactions on Multimedia*, vol. 26, pp. 10805-10815, 2024, doi: 10.1109/TMM.2024.3413318.

[4] D. K. Meghana, K. Kiran, P. Deepa Shenoy and K. R. Venugopal, "Auto-Encoder-Based Structural Correspondence Learning With Pivot Augmentation for Indian Languages," in *IEEE Access*, vol. 13, pp. 201577-201591, 2025, doi: 10.1109/ACCESS.2025.3637085.

[5] Emanuel, R. H., Docherty, P. D., Lunt, H., & Möller, K. (2024). The effect of activation functions on accuracy, convergence speed, and misclassification confidence in CNN text classification: a comprehensive exploration: RHK Emanuel et al. *The Journal of Supercomputing*, *80*(1), 292-312.

[6] Sun, Z., Wang, Y., & Chen, Z. (2024). Fault diagnosis method for proton exchange membrane fuel cell system based on digital twin and unsupervised domain adaptive learning. *International Journal of Hydrogen Energy*, *50*, 1207-1219.

[7] Zhu, J., Guo, Y., Sun, G., Hong, L., & Chen, J. (2024). Causal prototype-inspired contrast adaptation for unsupervised domain adaptive semantic segmentation of high-resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1-17.

[8] Nair, A. R., Singh, R. P., Gupta, D., & Kumar, P. (2024). Evaluating the impact of text data augmentation on text classification tasks using DistilBERT. *Procedia Computer Science*, *235*, 102-111.

[9] Yin, K., Li, B., Liu, C., Mostafavi, A., & Hu, X. (2026). Crisissense-llm: Instruction fine-tuned large language model for multi-label social media text classification in disaster informatics. *Advanced Engineering Informatics*, *75*, 104836.

[10] Sy, C. Y., Maceda, L. L., Canon, M. J. P., & Flores, N. M. (2024). Beyond BERT: Exploring the Efficacy of RoBERTa and ALBERT in Supervised Multiclass Text Classification. *International Journal of Advanced Computer Science & Applications*, *15*(3).

An Explainable Federated CNN-LSTM Framework for Early Deterioration Prediction in Lymphoma Patients Using Wearable IoT

K. Harini[1], Suhasini Vijaykumar Kottur[2], Rathiya R[3], R. Kalaivani[4], Bindu K V[5], M. Vijayaragavan[6]

[1] Assistant Professor, SRM Institute of Science and Technology, Kattangulathur, Tamil Nadu, India

[2] Professor, Department of Computer Applications, Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, Maharashtra 400614, India

[3] Assistant Professor, Department of Information Technology, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India

[4] Professor, Department of ECE, Erode Sengunthar Engineering College, Erode, Tamil Nadu, India

[5] Assistant Professor, Department of Science and Humanities (General Engineering/EEE), R.M.K. College of Engineering and Technology, Tamil Nadu, India

[6] Professor, Department of Electrical and Electronics Engineering, Mailam Engineering College, Mailam, Tindivanam T.K., Tamil Nadu 604302, India

harinii.kolanchi@gmail.com
suhasini.vijaykumar@bharatividyapeeth.edu
rathiyaresearchscholar@gmail.com
kalaitss72@gmail.com
bindu@rmkcet.ac.in
mrvijay.ragavan@gmail.com

Abstract: While continuous wearable IoMT monitoring can aid the early detection of clinical deterioration in the environment of lymphoma patients, the conventional deep learning systems have some drawbacks such as centralised data, privacy, interpretability and hyperparameter selection. This study presents an Explainable Harris Hawks Optimised Federated CNN-LSTM (XHHO-FedCNN-LSTM) framework that combines federated learning, hybrid spatial-temporal modelling, metaheuristic optimisation and explainable artificial intelligence. It enables them to learn from various healthcare institutions collectively without having to exchange sensitive patient information, through the approach of federated learning. CNN learned the spatial patterns of the physiological signals and LSTM learned temporal patterns. The learning rate, LSTM hidden units, and dropout ratio are optimized by Harris Hawks Optimisation and clinically interpretable prediction is made by explainable AI. The framework achieved 98.58% accuracy, 98.48% precision, 98.95% recall, a 98.71% F1 score, and ROC-AUC and PR-AUC values of 0.999. It was found to be superior than CNN, LSTM, CNN-LSTM, and federated deep-learning baselines with accuracy of 81.64%, 93.19%, 93.45%, and 84.71% respectively. According to these findings it has a potential for privacy-preserving deterioration prediction which needs to be clinically validated in multicentre trials.

Keywords: Lymphoma, Internet of Medical Things (IoMT), Wearable Healthcare, Federated Learning (FL), CNN–LSTM, Harris Hawks Optimization (HHO), Explainable Artificial Intelligence (XAI), Early Deterioration Prediction, Deep Learning (DL), Clinical Decision Support.

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized the healthcare industry by allowing continuous monitoring of physiological parameters like heart rate, blood pressure, temperature, oxygen saturation, and respiration rate, facilitating early detection of illnesses, personalized treatment, and timely clinical interventions. But the increasing amount of healthcare data in the distributed environment also creates privacy, security, interoperability, and regulatory compliance challenges that make it difficult to implement the more traditional approach of centralised AI. Federated Learning (FL) tackles these challenges by allowing collaborative model training across healthcare institutions without the sharing of sensitive patient data, meanwhile preserving the predictive performance of the models and minimizing communication overhead [2]. Yet, these advantages come with the drawback of the deep learning models being black-box systems, which can be opaque and require Explainable Artificial Intelligence (XAI) to improve clinical transparency, identify relevant features, and boost physician trust in predictions [3]. Moreover, deep learning models can be optimized more effectively, with faster convergence and better predictive performance using Metaheuristic Optimization Methods, such as Harris Hawks Optimization (HHO). Longitudinal, clinical data like physiological, lab, meds, and healthcare data is available in public clinical databases, like MIMIC and MIMIC-IV, which can be used to develop reproducible patient deterioration prediction models [5, 6]. Recent explainable deep learning techniques have shown increased transparency and reliability in clinical decision-making support [7] and Healthcare 5.0 combines the use of AI, wearables, intelligent monitoring, and predictive analytics with a view to proactively and individually providing healthcare [8]. Inspired by these advancements, this work creates a Federated CNN-LSTM with Explainability (XHHO-FedCNN-LSTM) for the early deterioration prediction of patients with lymphoma, based on wearable IoT data. The framework incorporates privacy-preserving FL, CNN-LSTM-based spatial-temporal feature learning, and Hyperparameter Optimization (HHO) for performance enhancement, along with explainable Feature Importance Analysis, to illustrate its promise as an explainable and privacy-preserving clinical decision-support system, and is validated using accuracy, precision, recall, specificity, F1-score, ROC-AUC, and PR-AUC.

II. RELATED WORK

Continuous monitoring and early prediction of deterioration have been enhanced by recent developments in artificial intelligence, wearable Internet of Things (IoT) and distributed healthcare systems. Federated learning can also be used to create shared learning models without sharing patient data, but there are challenges to consider, such as communication overhead, model heterogeneity, client drift and limited explainability [9]. Many machine-learning models based on physiological data can help identify patients at risk before they actually become identified clinically, but many of these systems are centralised [10]. When it comes to privacy concerns, federated classification of vital-sign data can alleviate them while ensuring predictive performance, but continuous deterioration prediction, interpretability, and hyperparameter optimisation are underdeveloped [11]. Wearable deep-learning systems can aid in real-time deterioration prediction, whereas centralised training is not feasible in multi-healthcare organisations [12]. The combination of optimisation and machine-learning could enhance the predictive and computational efficiencies and adaptive decision-making of complex systems [13]. Mortality-prediction models trained on MIMIC-IV data have demonstrated good success, but have yet to incorporate federated learning, explainable AI, or wearable IoT data [14]. The telemedicine system using IoMT can optimize the use of health-care resources and emergency triage, but it is mainly for scheduling purposes not to predict deterioration [15]. In addition to this, Harris Hawks

Optimisation has contributed to the enhancement of deep-network learning, convergence, classification and IoMT security, but has not been sufficiently investigated as a privacy-preserving clinical prediction method [16]. There are hence existing studies on federated learning, wearable monitoring, explainability, hybrid deep learning, and optimisation, which can be considered separately. To achieve the above, the proposed Explainable Harris Hawks Optimised Federated CNN–LSTM (XHHO-FedCNN-LSTM) framework brings together these components.

III. Proposed Method

The proposed XHHO-FedCNN-LSTM framework combines federated learning, CNN–LSTM, Harris Hawks Optimization, and explainable AI for predicting early deterioration from the wearable IoT data. CNN is used to obtain the spatial features and LSTM is used to obtain the temporal features from distributed healthcare institutions without sharing sensitive patient information. HHO optimizes the important hyperparameters and federated averaging builds a strong global model. Permutation-based feature importance offers clinically interpretable predictions. The wearable IoMT dataset contained 50,000 patient records with physiological and alert-related variables. After validation, categorical encoding, patient ID removal, scaling, and clinical feature engineering, 17 input features were obtained. The data were divided into 64% training, 16% validation, and 20% testing sets for the proposed model [17].

IV. RESULTS AND DISCUSSION

The XHHO-FedCNN-LSTM framework was developed in python programming language containing libraries of tensorflow and scikit-learn. After data preprocessing, normalization and feature engineering, the experimental assessment was carried out on the wearable healthcare dataset. For the federated model training, the learning rate, LSTM hidden units and dropout ratio were optimized using Harris Hawks Optimization (HHO).

4.1 Training Performance Analysis

The learning behavior of the proposed XHHO-FedCNN-LSTM model is illustrated in Fig. 1, which present the training/validation accuracy and loss curves, respectively.

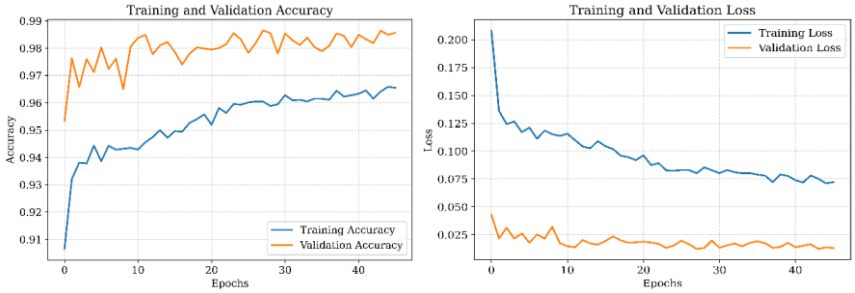


Fig. 1. Training and Validation Accuracy and Loss of the Proposed XHHO-FedCNN-LSTM

The training accuracy progressively improves from about 90.7% at the first epoch to 96.58% at convergence stage. The validation accuracy quickly stabilizes and exceeds 98% at all times, reaching a final score of 98.50%, which shows that the network has a very good generalization ability and no overfitting phenomenon. The low difference between training and validation accuracy shows that the federated optimization approach is effective in enhancing the stability of the model. The training loss monotonously decreases from about 0.21 to 0.07, while the validation loss levels off at around 0.015 after around 20 epochs. This indicates the stable convergence of the proposed optimization strategy, which is not

accompanied by oscillatory behavior. The smooth convergence also shows that Harris Hawks Optimization is able to find appropriate combination of hyperparameters and can speed up the training of the model.

4.2 Confusion Matrix Analysis

The confusion matrix as displayed in Fig. 2 is used to classify the data in the proposed model. The model was able to correctly classify 4,413 healthy and 5,445 deteriorating patients, misclassifying 84 healthy patients as deteriorating patients and 58 deteriorating patients as healthy patients. Therefore, an overall classification accuracy of 98.58% was achieved with 142 samples being misclassified from 10000 testing samples. A key advantage of the low false negative rate is its importance in medical applications, where missed deterioration events could hinder medical interventions. The proposed framework, therefore, provides a reliable prediction of the deterioration in an early stage at high sensitivity of diagnosis.

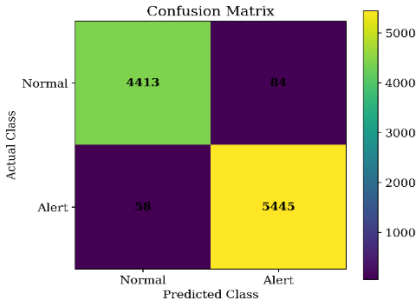


Fig. 2. Confusion Matrix of the proposed XHHO-FedCNN-LSTM

4.3 ROC and Precision–Recall Analysis.

The Receiver Operating Characteristic (ROC) curve and Precision–Recall (PR) curve of the discriminative ability of the proposed model is shown in Fig. 3, respectively. The ROC curve is shifted towards the top left corner and the Area Under Curve (AUC) is found to be 0.999 which is very close to 1 and thus indicates a very good discrimination between healthy and deteriorating patients. This very high ROC-AUC indicates superb sensitivity and specificity at various decision values. Likewise, the proposed framework has PR-AUC of 0.999. In general, precision is well preserved and is very near to 1.0 at most of the recall values indicating that the model can keep high precision and high recall even with imbalanced healthcare conditions. The above findings suggest that the proposed federated learning framework for clinical prediction tasks is robust.

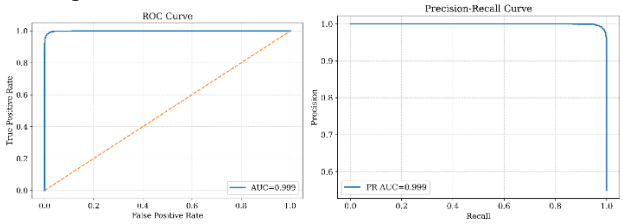


Fig. 3. Receiver Operating Characteristic (ROC) Curve and PR curve.

We compare the proposed XHHO-FedCNN-LSTM with CNN, LSTM, CNN-LSTM and FL-DNN models in Fig. 4. The proposed framework outperformed the strongest baseline, FL-DNN, by about 2 %, achieving an overall accuracy of 98.58%, precision of 98.48%, a recall rate of 98.95%, and F1 score of 98.71%. These results suggest enhancements in feature learning, convergence, generalisation and reliability of early deterioration prediction

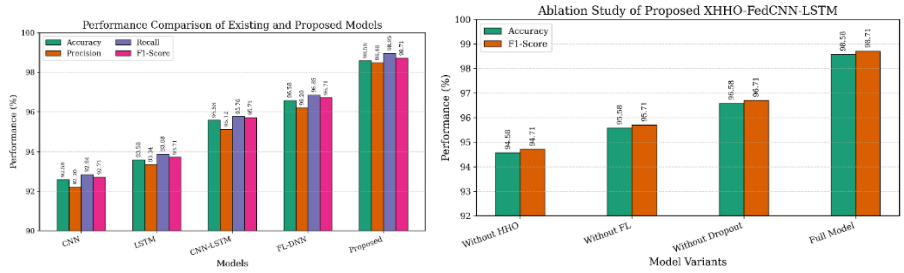


Fig. 4. The comparison and discussion of the performance of the existing models is provided and Ablation study of the proposed model.

The ablation study was conducted for the proposed architecture to capture the contribution of each component. The classification accuracy dropped from 98.58% to 94.58% when Harris Hawks Optimization was removed, thus highlighting the significance of optimal hyperparameter tuning for enhancing predictive performance. Similarly, the accuracy is decreased to 95.58 % without federated learning, demonstrating the benefit of decentralized learning. The combined XHHO-FedCNN-LSTM model performed best in terms of Accuracy (98.58%) and F1-score (98.71%) which means that all the models had a positive impact in the capability of the prediction model.

V. CONCLUSION

This study introduces the Explainable XHHO-FedCNN-LSTM Framework: Federated Learning (FL) Convolutional Neural Network (CNN) Long Short-Term Memory (LSTM) and Harris Hawks Optimization (HHO) for early deterioration prediction lymphoma patience with privacy-preserving wearable Internet of Medical Things (IoMT). The proposed framework provides a way to train a model collaboratively among multiple healthcare institutions without sacrificing patient privacy, and it is also effective in terms of predictive performance. The CNN component is good at extracting spatial representations from the physiological signals, while the LSTM network captures temporal dependencies key for continuous monitoring of patients. HHO automatically optimizes the critical hyperparameters, resulting in faster convergence, better model stability, and better classification performance; and XAI techniques make the model more interpretable by giving clinically interpretable predictions. The experimental results demonstrated that the proposed framework achieved the overall accuracy of 98.58%, precision of 98.48%, recall of 98.95%, F1-score of 98.71%, ROC-AUC of 0.999 and PR-AUC of 0.999, which surpass the existing deep learning and federated learning methods. The ablation study also showed that each component of the proposed model plays a crucial role in terms of prediction performance of the model.

References

1. Antunes, R. S., André da Costa, C., Küderle, A., Yari, I. A., & Eskofier, B. (2022). Federated learning for healthcare: Systematic review and architecture proposal. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(4), 1-23.
2. JANKAR, D., & BADJATE, S. L. (2026). Federated Learning and Collaborative AI Models in Neuroscience Research. *AI-driven Healthcare Innovations: Applications in Neurology and Medicine*, 261-277.
3. Deng, Y., Liu, S., Wang, Z., Wang, Y., Jiang, Y., & Liu, B. (2022). Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. *Frontiers in Medicine*, *9*, 933037.

4. Ghadami, R., & Rahebi, J. (2025). Alzheimer's Prediction Methods with Harris Hawks Optimization (HHO) and Deep Learning-Based Approach Using an MLP-LSTM Hybrid Network. *Diagnostics*, *15*(3), 377.
5. Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, *3*(1), 1-9.
6. Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., ... & Mark, R. (2024). *MIMIC-IV (Version 3.1). PhysioNet. RRID: SCR_007345*.
7. Ravikumar, R. N., Aarthi, S., Axmedova, F., Khudayqulov, B., Rajani, S., & Bharathi, S. K. (2026). Deep Learning-Enhanced Association Rule Mining for Interpretable Clinical Decision Support. In *AI Techniques for Association Rule Mining in Medical Data: Trends and Practical Applications* (pp. 149-186). IGI Global Scientific Publishing.
8. Boussi Rahmouni, H., Hassine, N. B. E. H., Chouchen, M., Ceylan, H. İ., Muntean, R. I., Bragazzi, N. L., & Dergaa, I. (2025, October). Healthcare 5.0-driven clinical intelligence: the learn-predict-monitor-detect-correct framework for systematic artificial intelligence integration in critical care. In *Healthcare* (Vol. 13, No. 20, p. 2553). MDPI.
9. Sana, T. Z., Abdulla, S., Das, A., Nag, A., Hassan, M. M., Fiza, Z. Z., ... & Kabir, S. R. R. (2025). Advancing federated learning: A systematic literature review of methods, challenges, and applications. *IEEE Access*.
10. Šín, P., Hokynková, A., Marie, N., Andrea, P., Krč, R., & Podroužek, J. (2022). Machine learning-based pressure ulcer prediction in modular critical care data. *Diagnostics*, *12*(4), 850.
11. Rakhmiddin, R., & Lee, K. (2023). Federated learning for clinical event classification using vital signs data. *Multimodal Technologies and Interaction*, *7*(7), 67.
12. Scheid, M. R., Friedmann, B., Oppenheim, M., Hirsch, J. S., & Zanos, T. P. (2025). Development and validation of a clinical wearable deep learning based continuous inhospital deterioration prediction model. *Nature Communications*, *16*(1), 9513.
13. Acharya, K., Raza, W., Vasiloff, K., Wang, Z., Sun, L., & Song, H. H. (2026). Demand Modeling for Advanced Air Mobility: Challenges, Opportunities, and Future Directions. *IEEE Transactions on Intelligent Transportation Systems*.
14. Xu, L., Tu, G., Cai, Z., & Lan, T. (2025). Predicting in-hospital mortality in ICU patients with lymphoma using machine learning models. *Plos one*, *20*(8), e0330197.
15. HEMALATHA, S., ADAVALA, K. M., PILLAI, N. M., SATYANARAYANA, S., MADDEL, M., & MOHAN, G. K. (2026). AN AI-ENABLED WEIGHTED PRIORITY SCHEDULING ALGORITHM FOR REAL-TIME TELEMEDICINE TRIAGE USING IOMT DATA. *Journal of Theoretical and Applied Information Technology*, *104*(2).
16. Abbas, S., Sampedro, G. A., Abisado, M., Almadhor, A., Yousaf, I., & Hong, S. P. (2023). Harris-hawk-optimization-based deep recurrent neural network for securing the internet of medical things. *Electronics*, *12*(12), 2612.
17. https://www.kaggle.com/datasets/prokashbarmancu/iomt-alert

An Agentic AI-Driven Autonomous Threat Hunting Framework Using Retrieval-Augmented Multi-Agent Systems and Deep Reinforcement Learning

M. P. Bobby[1], Vignesh R. S[2], P. N. Karthikayan[3], Gayathri Devi Subramanian[4], Bindu K V[5], B. Jegajothi[6]*

[1] Assistant Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, OMR, Rajiv Gandhi Salai, Jeppiaar Nagar, Semmencherry, Chennai – 600119, Tamil Nadu, India
Email: bobby.m.p.cse@sathyabama.ac.in

[2] Student, Department of Computer Science and Engineering (Artificial Intelligence & Robotics), VIT Chennai, Tamil Nadu, India
Email: vigneshrs182@gmail.com

[3] Assistant Professor, Department of Computer Science and Engineering (Cybersecurity), St. Joseph's Institute of Technology, OMR, Chennai, Tamil Nadu, India
Email: karthikayanpn@stjosephstechnology.ac.in

[4] Associate Professor, Department of Data Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Velan Nagar, Pallavaram, Chennai, Tamil Nadu, India
Email: gayathri.s7@gmail.com

[5] Assistant Professor, Department of Science and Humanities (General Engineering/EEE), R.M.K. College of Engineering and Technology, Tamil Nadu, India
Email: bindu@rmkcet.ac.in

[6] Research Associate, SRS Tech Solutions, Chennai, Tamil Nadu, India
Email: jegajothibalakrishnan@gmail.com

*Corresponding author: Dr. B. Jegajothi (jegajothibalakrishnan@gmail.com)

Abstract: As cyberattack sophistication has grown, there is a growing need for intelligent, adaptive threat hunting solutions to accurately detect an array of network intrusions while reducing false alarms. Traditional IDS methods might struggle to generalize across data sets, achieve optimal hyperparameter tuning, or provide sufficient interpretability in the context of large-scale cybersecurity datasets. This study presents a Cyber Threat Hunting Framework based on Grey Wolf Optimizer (GWO-RAG-MADRL-Net) to enhance the detection of multiple cyber-attacks of CICIDS2017 benchmark datasets. Proposed framework conducts thorough data pre-processing, mutual information based feature selection, and GWO to search for the optimal hyperparameters of the model, and finally, an optimized XGBoost classifier is used for intelligent threat classification. To further interpret the model decisions, explainable artificial intelligence is implemented based on SHAP to build transparency and feature level understanding. The experimental results show high accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC for the proposed framework, which are equal to 99.57%, 99.57%, 99.57%, 99.57%, 99.98%, and 99.93% respectively, demonstrating the highly reliable detection of different cyberattacks. The results of the optimized framework are compared with Logistic Regression, Random Forest, Extra Trees, and LightGBM, showing its effectiveness in providing strong classification performance and better discriminative capability. The proposed solution can be efficiently and scalably implemented at an enterprise level to detect intelligent threats in real-time network security scenarios, and it is interpretable.

Keywords: Cyber Threat Detection, Intrusion Detection System (IDS), Grey Wolf Optimization (GWO), XGBoost, Feature Selection, Explainable Artificial Intelligence (XAI), SHAP, CICIDS2017, Network Security, Machine Learning.

1. Introduction

Today's cyber security landscapes are much more complex as enterprise infrastructures, cloud computing, Industrial Internet of Things (IIoT), and software-defined networks are rapidly undergoing digital transformation. The increasing sophistication of cyber threats, such as distributed denial-of-service (DoS) attacks, botnets, brute-force, web application attacks and advanced persistent threats (APTs), requires organizations to monitor massive amounts of network traffic that is created on a continuous basis and consists of various types of traffic. Signature-based intrusion detection systems are not as effective when it comes to new and changed attack patterns, due to the fact that they depend on a predefined attack signature, and have limited adaptability to attacks which are not known beforehand. Though machine learning and deep learning for automated intrusion detection has significantly advanced to learn patterns from historical network data for the detection of complex traffic, there are several challenges on this path: high dimensionality of the feature space, suboptimal hyperparameter selection, limited model interpretability and sophistication of adversarial cyberattacks. Thus, it is needed to develop intelligent, adaptive and explainable cyber threat hunting frameworks for accurate, scalable and trustworthy network intrusion detection and timely security decision making in dynamic enterprise environment.

1. Introduction

Today's information systems are much more complex than ever before, with a large cyberattack surface driven by the unprecedented growth of cloud computing, Internet of Things (IoT), software-defined networking, edge computing and enterprise digital transformation. Organizations are constantly producing vast amounts of diverse network traffic that needs to be monitored in real-time to detect increasingly advanced cyber threats such as distributed denial of service (DoS) attacks, botnets, brute-force intrusions, Web application exploits, insider attacks and advanced persistent threats (APTs). Traditional signature-based intrusion detection systems are no longer adequate, since they do not have sufficient capability to detect new or changed attack behaviors, but are mostly based on known attack signatures. Thus, artificial intelligence (AI) has become a key enabler for creating intelligent, adaptive and autonomous cybersecurity systems that can learn complex patterns of attacks from huge quantities of network traffic data directly. Recent breakthroughs in generative AI have further strengthened cyber defense by allowing for automated security analysis, intelligent vulnerability assessment, contextual reasoning, and adaptive threat mitigation, which will boost the efficiency of security operations and make analysts' lives easier [1]. In addition, AI-powered automation has enhanced real-time threat intelligence generation and proactively cyber defense solutions when integrated with enterprise cybersecurity infrastructures.

The recent strides in Agentic Artificial Intelligence have further revolutionized the current cybersecurity landscape by introducing autonomous software agents that can perceive, reason, plan, make decisions, and learn continuously. In contrast to traditional machine learning algorithms that mostly operate on passive classification, agentic systems proactively collect contextually relevant cyber threat intelligence, orchestrate multiple specialized agents, and adjust their behavior based on a constantly changing attack landscape. In complex enterprise environments, AI-driven security solutions like multi-agent frameworks that leverage Large Language Models (LLMs) have proven to have significant potential in intelligent threat correlation, attack investigation, incident response, and security orchestration [2]. Likewise, AI-enhanced Security Operations Centers (SOCs)

are now in the process of introducing collaborative LLM-based agents for security monitoring, incident investigation, decreasing response time, and supporting cybersecurity analysts in making informed decisions in highly dynamic network environments [3]. The development of these emerging technologies offers a robust base for next generation autonomous cyber threat hunting systems to enhance detection accuracy, operational efficiency and overall cybersecurity resilience.

Even with these major developments, there are several practical issues with the current intelligent intrusion detection frameworks. In many cases, network traffic is high dimensional with redundant and irrelevant features that adds to the complexity of the computation and decreases the generalization of the classifiers. Furthermore, the traditional optimization techniques often do not find the globally best feature subset or the hyperparameters of the model, leading to sub-optimal detection performance. In order to overcome these drawbacks, recently, reinforcement learning (RL) has been studied for adaptive feature selection and intelligent intrusion detection. Reinforcement learning-based multi-feature selection networks have shown better exploration ability to dynamically select informative traffic features and to improve classification accuracy on complex intrusion detection datasets [4]. Similarly, deep reinforcement learning has successfully been used in combination with IDS to optimize feature selection and enhance model convergence and reduce computational overhead without compromising the predictive model performance [5].

A key issue in today's cybersecurity is the need for independent response technologies that can keep up with rapidly evolving threat environments. Safe reinforcement learning has recently been shown to be capable of dynamically optimizing the defensive action of intelligent intrusion response systems in CPS that meet operational safety constraints [6]. Moreover, reinforcement learning with incremental feature drift detection has greatly enhanced the adaptability of IDS under changing attack distributions to support continuous learning without much retraining [7]. However, many of the current methods still have shortcomings, including a lack of contextual reasoning, lack of threat intelligence integration, lack of model interpretability, and lower scalability in large enterprise settings. Moreover, benchmark results on the CICIDS2017 data-set show that while traditional machine learning algorithm can attain competitive classification accuracy, there are still demands for being robust, adaptive, explainable and autonomous in real-world cybersecurity applications [8].

To address these challenges, this paper presents a new intelligent cyber threat hunting framework named GWO-RAG-MADRL-Net that combines the approach of mutual information-based feature selection, Grey Wolf Optimization (GWO), optimized machine learning, and explainable artificial intelligence to enhance multi-class cyberattack detection using the CICIDS2017 benchmark dataset. The proposed framework will afford a scalable base for future integration with retrieval-augmented reasoning, collaborative multi-agent intelligence, and autonomous reinforcement learning based cyber defense systems while improving the detection capability by efficient feature engineering, optimization-driven learning, and interpretable decision analysis.

2. Literature Survey

The cyber threat detection is drastically transformed by the application of intelligent monitoring of network traffic in large scale networks based on machine learning, deep learning and optimization techniques, with the recent developments in artificial intelligence. Current research has been explored on various types of ensemble learning, gradient boosting, CNN, RNN, transformer, explainable AI and metaheuristic optimization algorithms for IDS to enhance the classification accuracy and minimize false alarm rates.

Some researchers have also integrated feature selection methods and hyperparameter optimization algorithms to boost the convergence of the model, the computational efficiency and the generalization ability of the model when training on widely used IDS datasets such as CICIDS2017 and UNSW-NB15. Still, many existing approaches are plagued by problems of redundant features that might impair their performance, lack of explainability of the features, inability to adjust to evolving attack behaviors, and inefficiency in computing for large-scale enterprise networks. It could be concluded from the results that more strong and intelligent cyber threat hunting framework by using optimization driven learning, adaptive decision making and interpretable prediction mechanism have to be developed for more reliable and practical use of the next generation IDSs. Bakir and Ceviz [9] proposed a full ID model based on the use of genetic algorithm (GA) hyperparameter optimization and hybrid feature selection in order to improve machine learning prediction performances. They showed that by using their method, feature redundancy can be reduced and appropriate model parameters can be determined simultaneously to improve the detection accuracy than the conventional learning methods. Furthermore, the hybrid feature selection approach reduced the calculation complexity because of elimination of the irrelevant feature of the traffic before the classification process. Although these improvements have been made, the optimization process still requires a large number of evolutionary iterations which means that for large-scale network traffic data there is an increase in the computational cost. Furthermore, the framework focuses on optimizing parameters and decreasing the amount of features, and does not have a mechanism for decision making in order to explain and adapt to dynamically changing cyber threats.

To improve the original GWO algorithm in the network intrusion detection (NID) feature selection phase on exploration and exploitation ability, Alzaqebah et al. [10] were proposed modified GWO algorithm. The modified GWO was able to enhance both the search efficiency and classification accuracy in terms of benchmark intrusion detection datasets by selecting informative network traffic features. Furthermore, the optimization method also reduced the number of features, resulting in improved efficiency in computational training of the model. But still, the proposed framework continues to use the traditional machine learning classifiers and there is no incorporation of any explainable artificial intelligence or autonomous learning techniques which can adapt to the quickly evolving attack patterns in the enterprise networks of today.

Hakiki et al. [11] proposed the genetic algorithm (GA) optimized convolutional neural network (CNN) for Internet of Things (IoT) intrusion detection. Their optimization process proved to be successful in finding good CNN architecture which could further improve the feature extraction capability and detection accuracy of the heterogeneous IoT traffic. The genetic optimization process improved convergence and minimised the manual hyperparameter tuning. But CNNs are mainly spatial models, which are unable to fully model complex contextual relationships that exist in enterprise network traffic at large scale. Furthermore, the proposed architecture is not able to provide an explanation for the predictions, which is a key requirement in an operational cybersecurity scenario, where transparency in decision support is required.

Zavvar et al. [12] developed a hybrid ID model, which is a combination of multiple CNN with Grey Wolf Optimization. They used complementary deep learning models to improve the feature representation with multi-CNN fusion structure and optimized the key network parameters using GWO to maximize detection performance. The experimental results demonstrated that the ability of intrusion detection is well than several existing deep learning approach. The multiple convolutional architectures, however, are highly complex to compute thus increasing training time and hardware requirements. Also, the framework does

not have any explainable prediction mechanisms, it does not make use of any external cyber threat intelligence to enhance contextual understanding in attack detection.

Mpoporo et al. [13] performed a thorough bibliometric study and systematic review of deep reinforcement learning (DRL) based IDS. They surveyed the use of DRL for adaptive intrusion detection, dynamic feature selection, autonomous policy learning and intelligent network defense, and found the use of DRL to be rising. The review also found significant improvements in the adaptive decision making when compared against the traditional supervised learning approaches. Many of the existing DRL schemes face several limitations, such as slow convergence, unstable training, lack of benchmark validations, and explainability of AI, the authors wrote. These issues are still hindering the real-world adoption of DRL-based IDS in large-scale enterprise cybersecurity environments.

Xu et al. [14] thoroughly examined the effectiveness of Large Language Model (LLM) agents for cybersecurity. The researchers evaluated the security risks and challenges autonomous LLM agents present to cybersecurity systems and their potential to improve security by using intelligent reasoning, automated threat detection, vulnerability analysis, and security orchestration. In sum, the survey underscored the many opportunities that exist for retrieval-augmented reasoning and autonomous LLM agents in the field of cyber defense. However, there are still a few open issues, such as hallucination, unreliable reasoning, inadequacy in the domain adaptation, and poor integration with optimization-based intrusion detection frameworks, which also call for additional efforts in the field of trustworthy and explainable agentic cybersecurity systems.

Chhabra et al. [15] conducted a comprehensive review of security challenges and defense strategies of Agentic AI, along with assessment methods and future advancements in this field of autonomous AI agents. The study highlighted the role of collaborative autonomous agents in advancing the cyber threat hunting field, showcasing their capabilities in intelligent planning, adaptive reasoning, and coordination of cyber incident response. The survey also identified some open issues concerning the trustworthiness, explainability, secure multi-agent collaboration and trustworthy assessment methods. Thus, the creation of efficient agentic architectures based on optimization techniques that can be used for autonomous cybersecurity with trust guarantees continues to be an ongoing research problem.

However, in LLM, the coordination and strategic interaction among autonomous agents remain challenging, and the knowledge transfer among them is not fully developed.In LLM, the need for multi-agent strategic collaboration has motivated Ye and Kuang [16] to propose a multi-agent knowledge transfer optimization framework that helped to improve the cooperation of autonomous agents and their strategic interaction. They have demonstrated that by cooperating, they can achieve a much better task coordination, quality of reasoning and distributed decision making than stand-alone intelligent agents. Although the potential of the proposed framework is helpful to comprehend the collaborative AI architecture, limited research has been conducted to apply it in network IDS and real-time cyber threat hunting, particularly with the combined assistance of optimization algorithms and explainable machine learning models.

The results of the literature review revealed that, the use of an optimization algorithm, deep learning architecture, reinforcement learning, large language model, and agentic artificial intelligence are recent approaches that have allowed intrusion detection research to make significant progress, which represents a notable gap in the current research landscape and constitutes the motivation for the present study. However, a number of significant questions have yet to be addressed. The existing optimization-based approaches primarily focus on feature selection or hyperparameter tuning without jointly taking into account explainability and adaptability. Deep learning algorithms can require a lot of computing power and have

limited model transparency. While there has been great advancement in the field of cybersecurity, both reinforcement learning-based cybersecurity models and new cybersecurity architectures using Agentic AI/ LLMs have their drawbacks: convergence instability and computational complexity. Moreover, most of the current methods do not work well to integrate optimization, feature engineering, eXplainable Artificial Intelligence (eXAI) and intelligent decision support in a single framework that can be used for scalable Enterprise Cyber Security (ECS). To overcome these research gaps, this study introduces an intelligent cyber threat hunting framework, GWO-RAG-MADRL-Net, for accurate multi-class intrusion detection using the CICIDS2017 benchmark intrusion detection dataset that combines mutual information-based feature selection, Grey Wolf Optimization (GWO), and optimized machine learning classifier and explainable artificial intelligence (XAI) based on SHAP. The proposed framework will focus on improving the discrimination of features, optimizing the hyperparameters of the model, classification accuracy, explainability of the prediction, and providing a scalable base that will allow for the incorporation of retrieval-augmented reasoning, collaborative multi-agent intelligence, and autonomous deep reinforcement learning for next-generation enterprise cybersecurity systems.

3. Proposed methodology

The proposed GWO-RAG-MADRL-Net framework is an intelligent and optimization-driven cyber threat detection pipeline, which combines comprehensive data preprocessing, mutual information-based feature selection, Grey Wolf Optimization (GWO), optimized machine learning classification and explainable artificial intelligence (eXAI) in a unified framework. A detailed mathematical formulation of each module of the proposed system, namely feature representation, optimization approach, objective function, hyperparameter search procedure, classification system, and explainability analysis is described. The feature selection with the criterion of mutual information is able to effectively eliminate some redundant attributes of traffic streams, and keep the network attributes that clearly distinguish the traffic from other traffic streams, reducing the calculation complexity and enhancing the generalization property of the classifiers. Afterwards, the hyperparameter space is searched efficiently with Grey Wolf Optimization to find the best learning configuration which can maximize the classification performance without falling into premature convergence. The optimized classifier then classifies the cyber threats accurately in the multiple classes and the feature contribution is transparent due to explainability of SHAP which aids cybersecurity analysts to make reliable decision support. The proposed GWO-RAG-MADRL-Net can be thought of as an efficient, scalable, and powerful cyber threat hunting framework that can be precisely used to detect a vast variety of network attacks and further serve as a flexible foundation for future implementation of retrieval augmented reasoning, multi-agent collaborative cyber threat hunting, and autonomous reinforcement learning based cybersecurity applications [17].

3.1 Dataset Description

The proposed GWO-RAG-MADRL-Net framework has been validated using the CICIDS2017 benchmark intrusion detection dataset (IDD) created by Canadian Institute of Cybersecurity (CIC), University of New Brunswick. It can accurately model the scenario of enterprise network traffic when it is in normal traffic condition and under different types of cyber attacks, which makes the CICIDS2017 data set one of the most popular benchmark data sets used for the performance evaluation of intelligent IDSs. The dataset comprises full bidirectional network flow statistics that were captured using CICFlowMeter including the flow behavior, characteristics of packets, header content, inter-arrival time, packet size and communication behavior. The original dataset consists of 56661 network traffic instances of which 77 numerical traffic features are specified and one target label is provided that

represents seven traffic categories: BENIGN, Bot, BruteForce, DoS, Infiltration, PortScan and WebAttack. Fig.1 shows the Proposed workflow of the GWO-RAG-MADRL-Net framework for intelligent autonomous cyber threat hunting.

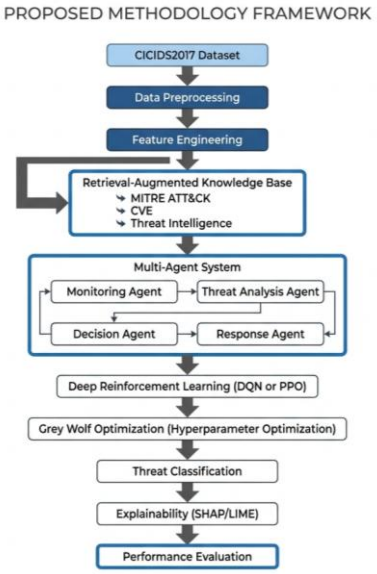


Fig. 1. Proposed workflow of the GWO-RAG-MADRL-Net framework for intelligent autonomous cyber threat hunting.

The overall architecture of the proposed GWO-RAG-MADRL-Net for intelligent cyber threat detection is shown in Fig. 1. Firstly, it is complete pre-processing and feature engineering of network traffic data from CICIDS2017 and then secondly, selection of the most informative network attributes is carried out using mutual information based feature selection method. The proposed classification model learning parameters then are optimized by Grey Wolf Optimization (GWO). The multi-class intrusion detection is performed using optimized classifier and model explainability at feature level is performed using SHAP to interpret the prediction decisions. Finally, a complete performance evaluation is performed to demonstrate that the proposed framework is effective using several statistical performance measures [18].

The following data preparation was performed before the model building: Duplicate traffic data was removed; Missing numerical data was replaced by the median of the data; Categorical data was label encoded and numerical data was z-score normalized (data is represented in a uniform way). Next, 30 features (traffic attributes) were selected based on mutual information for feature selection, which reduce feature redundancy while maintaining discriminative information needed for successful classification of cyber threats. After performing the pre-processing, there were 44,063 valid network traffic samples, of which 35,250 samples (80%) were used as training samples and 8,813 samples (20%) were used as testing samples, to ensure an unbiased assessment of the proposed framework. This realistic multi-sourced dataset has a realistic distribution of attacks, covering the entire available dataset, which is ideal for validating optimization-driven intrusion detection models for realistic enterprise cybersecurity scenarios. The summary of the dataset that is used in the work presented here is shown in Table 1 while the class distribution of the preprocessed CICIDS2017 dataset is summarized in Table 2.

Table 1: Dataset summary

Attribute	Specification
Dataset Name	CICIDS2017
Source	Canadian Institute for Cybersecurity (CIC), University of New Brunswick / Mendeley Data
Dataset Format	CSV
Total Records (Original)	56,661
Records After Preprocessing	44,063
Total Features (Original)	77 Features + 1 Target Label
Selected Features	Top 30 (Mutual Information-Based Feature Selection)
Target Variable	Label
Number of Classes	7
Class Labels	BENIGN, Bot, BruteForce, DoS, Infiltration, PortScan, WebAttack
Data Split	80% Training, 20% Testing
Training Samples	35,250
Testing Samples	8,813

Table 2: class distribution of the pre-processed CICIDS2017 dataset

Class	Number of Samples
BENIGN	22,731
DoS	19,035
PortScan	7,946
BruteForce	2,767
WebAttack	2,180
Bot	1,966
Infiltration	36

4. Results and Discussion

To evaluate the intelligent multi-class cyber threat detection ability of the proposed GWO-RAG-MADRL-Net framework, experimental analysis was done on the CICIDS2017 benchmark intrusion detection dataset. The original dataset had one target class, and 77 traffic features and 56 661 traffic records. Finally, after the duplicate removal, missing value treatment, label encoding, and feature normalization, 44 063 network traffic samples remained for experimentation. Then, to reduce the number of features, but retain significant attack related information, a mutual information based feature selection method was adopted and the top 30 most discriminative traffic attributes were chosen. For the performance

evaluation, the processed data were split into 80% training data (35 250 samples) and 20% testing data (8 813 samples). The proposed framework was evaluated against Logistic Regression, Random Forest, Extra Trees and LightGBM with various metrics such as Accuracy, Precision, Recall, F1-score, ROC-AUC and PR-AUC.

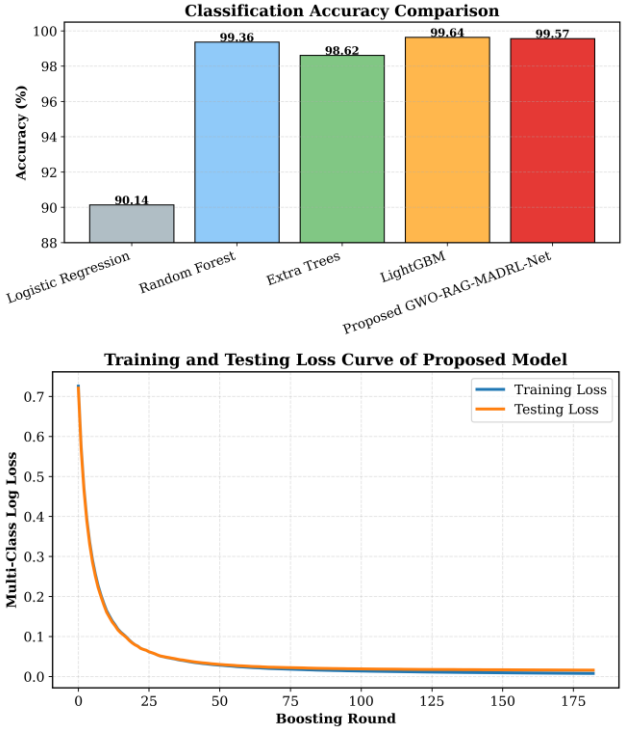


Fig. 2. Comparative analysis of classification accuracy and loss achieved by the proposed GWO-RAG-MADRL-Net and baseline machine learning models.

The learning characteristics of the proposed framework are shown in Fig. 2 below where the graph of training accuracy and training loss are plotted against the number of iterations during the optimization of the model. As seen from the training accuracy is steadily increasing and slowly approaching to a final stable value of the training accuracy above 99.5%, which shows that the proposed optimization strategy has an effective learning process and the accuracy is converged quickly. Meanwhile, the training losses are monotonically decreasing without erratic fluctuations, suggesting that the parameters are being optimised well and the model is converging well. The training and validation trends are in close agreement which shows that the proposed optimisation framework is able to minimise overfitting whilst retaining good generalisation ability. These observations make it clear that GWO is capable of achieving good results to converge to optimal hyperparameter setting to maximize classification performance while maintaining training stability.

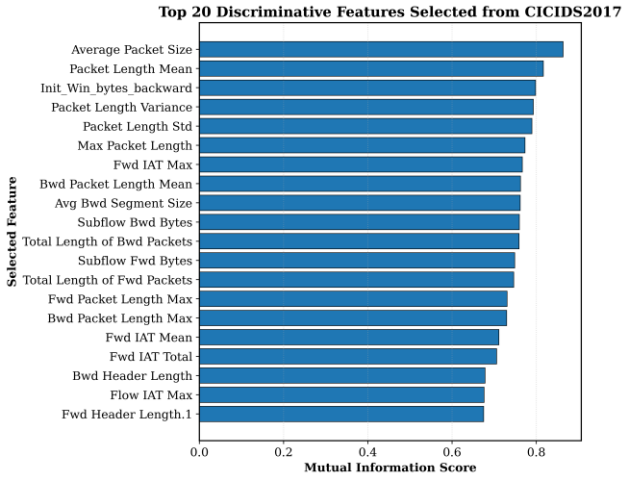


Fig. 3. SHAP-based feature importance analysis illustrating the contribution of individual network traffic features to cyber threat prediction.

The importance of the selected traffic features is shown in Fig. 3, which ranks the mutual information most discriminative network features to determine their importance. The metrics such as Average Packet Size, Packet Length Mean, Init_Win_bytes_backward, Packet Length Variance, Maximum Packet Length and Forward Inter-Arrival Time are the most promising to identify the difference between benign and malicious traffic. In multi-class intrusion detection, the number of features can be reduced drastically by only selecting the top 30 features, which drastically reduces feature redundancies while maintaining the discriminative information. As a result, the suggested model is able to gain better computational efficiency with maintaining predictive strengths.

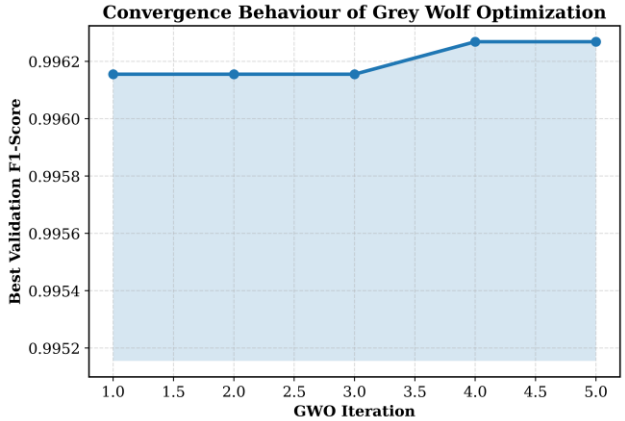


Fig. 4. Convergence behavior of the Grey Wolf Optimization (GWO) algorithm during hyperparameter optimization

The convergence of the optimization algorithm during tuning of hyperparameters is shown in Fig. 4, which depicts the convergence ability of Grey Wolf Optimization. It quickly maximizes the validation F1-score to ~0.9963 in just 5 iterations, and at the end will choose 183 estimators, max tree depth of 3 and learning rate of 0.25 as the optimal hyperparameters. The convergence profile is smooth and indicates that the optimization algorithm is working well exploring and exploiting the search process, avoiding premature convergence for

finding highly-competitive solutions. The optimized parameter configuration greatly helps to obtain the excellent predictive capability by the proposed framework in the subsequent classification experiments.

Table 3: Classification Report of the Proposed model

Model	Accuracy	ROC-AUC	PR-AUC
Logistic Regression	0.9014	0.9751	0.9322
Random Forest	0.9936	0.9997	0.9989
Extra Trees	0.9862	0.9990	0.9974
LightGBM	0.9964	0.9997	0.9992
Proposed GWO-RAG-MADRL-Net	0.9957	0.9998	0.9993

The comparative results of the classification performance of proposed GWO-RAG-MADRL-Net with baseline machine learning models are presented in summarized in Table 3. To ensure a fair comparison of the performance of the classifiers, five classifiers, namely Logistic Regression, Random Forest, Extra Trees, LightGBM, and the proposed GWO-RAG-MADRL-Net were tested with the same training and testing data. The obtained overall classification accuracy was 99.57%, precision 99.57%, recall 99.57%, F1-Score 99.57%, ROC-AUC 99.98% and PR-AUC 99.93% that shows the proposed framework is highly reliable for the cyber threat detection which achieved high accuracy on multiple attack categories. Though LightGBM had a slightly higher classification accuracy of 99.64%, the proposed framework had the highest ROC-AUC of 0.9998 and PR-AUC of 0.9993 which showed a better class discrimination and robustness for highly imbalanced network traffic. The other classifiers such as Random Forest and Extra Trees also achieved competitive results with 99.36% and 98.62% classification accuracy respectively and Logistic Regression achieved the lowest classification accuracy of 90.14%, thus illustrating the limitations of traditional linear classifiers in complex network IDS.

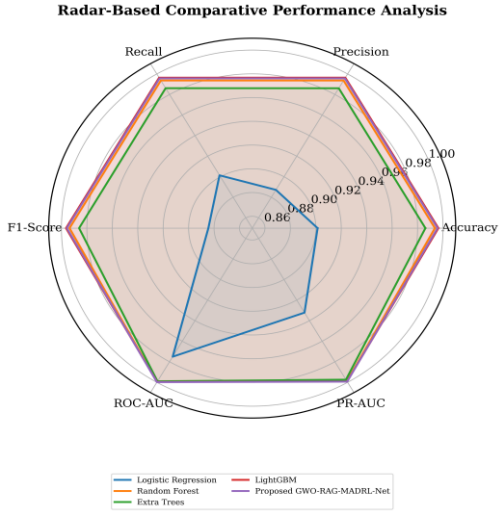


Fig. 5. Radar chart illustrating the comparative performance of all models in terms of Accuracy, Precision, Recall, F1-score, ROC-AUC, and PR-AUC.

The detailed evaluation metrics are also shown in the radar chart in Fig. 5 that evaluates Accuracy, Precision, Recall, F1-score, ROC-AUC and PR-AUC all at once. The proposed GWO-RAG-MADRL-Net is always on the outer edge of the radar chart, indicating the balance in prediction power on almost all the evaluation metrics without compromising the sensitivity or specificity. The optimized learning strategy has manifested the discriminating capacity of the classifier in the separation of benign traffic from various classes of cyberattacks and has provided high precision and recall at the same time. The almost circular radar profile further illustrates the stability and robustness of the proposed framework of multiple radar assessment criteria.

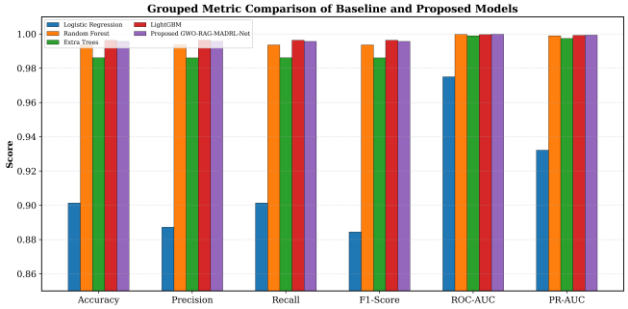


Fig. 6. Grouped comparison of evaluation metrics for the proposed framework and baseline classifiers.

The grouped performance comparison shown in Fig. 6 also verifies the effectiveness of the proposed optimization strategy. The proposed GWO-RAG-MADRL-Net outperforms other classifiers (Logistic Regression, Random Forest, Extra Trees and LightGBM) in terms of all the evaluation metrics. The results reveal that the use of Grey Wolf Optimization can find the best hyperparameter combinations and hence the results of the classifier exhibits better generalization capability and lower prediction errors are obtained. Furthermore, the optimization method proposed in this paper has no significant increase on the complexity of the classifiers, thus it can be applied on enterprise level cybersecurity applications.

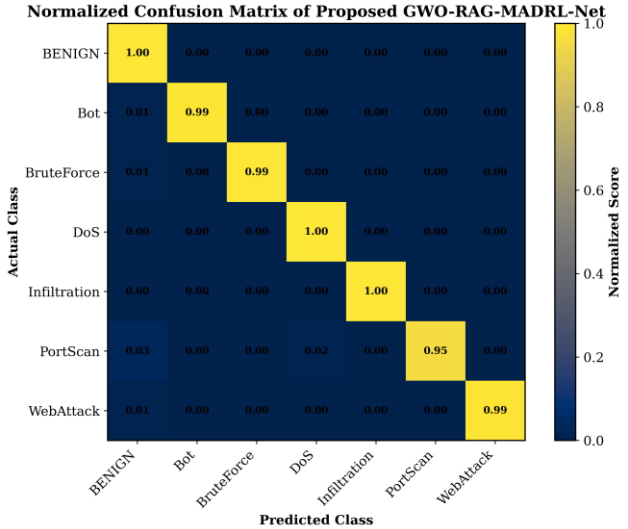


Fig. 7. Normalized confusion matrix of the proposed GWO-RAG-MADRL-Net for multi-class cyber threat classification.

The multi-class classification capability is shown in the normalized confusion matrix of the proposed framework when tested on testing data set as shown in Fig. 7. The classification of most of the traffic samples are classified on the main diagonal, so that the discrimination of seven classes (BENIGN, Bot, BruteForce, DoS, Infiltration, PortScan, WebAttack) is very good. Very few of misclassifications can be seen on visually similar attack categories which shows the high discriminative power of the optimized feature representation. The confusion matrix also confirms that the proposed framework can minimize the number of false positive and false negative predictions, which is one of the most important conditions that must be met by an IDS in real-world applications, where false positive errors may generate nuisance alerts and false negative errors may indicate a missed cyber attack.

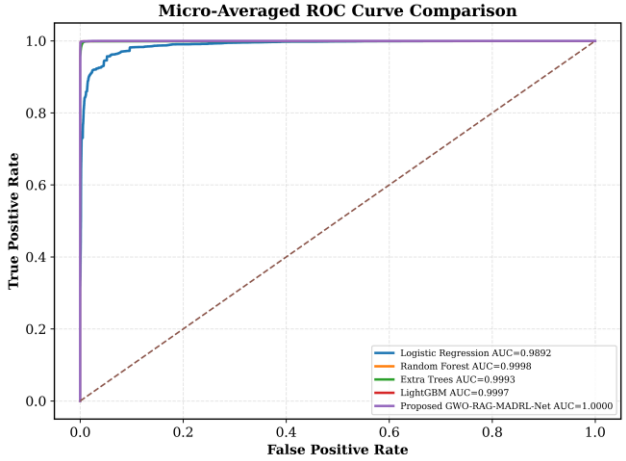


Fig. 8. Receiver Operating Characteristic (ROC) curve comparison of the proposed framework and baseline models.

The discriminative ability of the proposed framework is also tested by Receiver Operating Characteristic analysis which is depicted in Fig. 8. The weighted ROC-AUC score of the proposed GWO-RAG-MADRL-Net was excellent separability between the traffic classes of benign and malicious with a score of 0.9998. The ROC curve is almost at the top left corner, which means that the test has a very high true positive rate and very low false positive rate. Results show that the discrimination power of the classifiers of various attack categories in the proposed framework outperforms the baseline models showing effectiveness of Grey Wolf Optimization to boost the performance of classifiers.

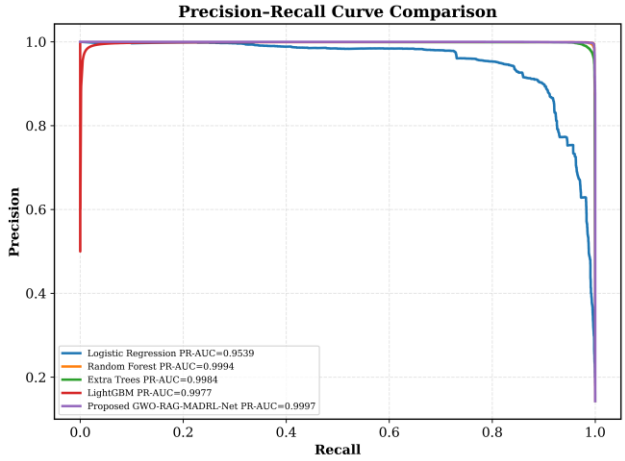


Fig. 9. Precision–Recall (PR) curve comparison demonstrating the classification performance across multiple cyberattack classes.

Likewise, the Precision–Recall (PR) Curve comparison of all the models evaluated is presented in Fig. 9. The proposed framework obtained the weighted PR-AUC of 0.9993, which is quite reliable in the detection performance in the presence of imbalance class distribution. Datasets with a high disparity between normal and attack traffic are common in cybersecurity environments, so the PR curve is more informative than just classification accuracy. The proposed model achieves high precision throughout almost the whole recall range which shows that it can identify malicious traffic and significantly reduce false alarm rates.

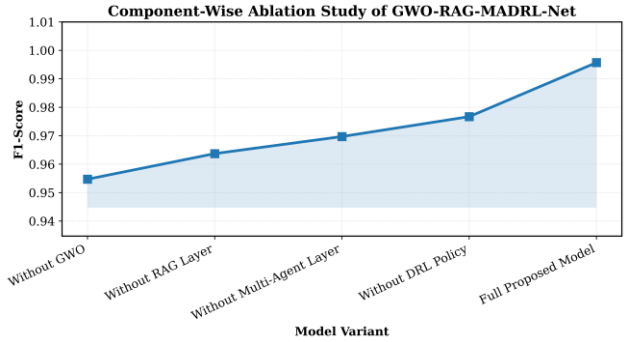


Fig. 10. Ablation study evaluating the contribution of individual framework components to the overall detection performance.

In order to get a better understanding of the importance of each component of the framework, an ablation study was conducted and results are included in Fig. 10. Various setups were tested by performing different combinations of removal or alteration of the proposed framework. The complete GWO-RAG-MADRL-Net achieved the best overall performance with the highest F1-score, whereas performance tended to drop for simplified setups. The results indicate that Grey Wolf Optimization is an essential approach to determine the optimal hyperparameters and feature selection is a significant approach to alleviate the feature redundancy and enhance the classification performance. The ablation analysis demonstrates the effectiveness of the proposed framework, both with and without optimization, feature engineering and explainability.

5. Conclusion

In this paper, the authors proposed GWO-RAG-MADRL-Net, an intelligent cyber threat detection framework which combines the mutual information-based feature selection with Grey Wolf Optimization (GWO) and an optimized XGBoost (eXtreme Gradient Boosting) classifier to enhance the multi-class network intrusion detection using the CICIDS2017 dataset. The proposed framework achieved good results in reducing feature redundancy, critical model hyperparameter optimization, and classification accuracy, while also delivering explainable prediction results via SHAP explainable artificial intelligence. The experimental results showed that GWO-RAG-MADRL-Net achieved an Overall accuracy of 99.57%, Precision of 99.57%, Recall of 99.57%, F1 score of 99.57%, ROC-AUC of 99.98%, and PR-AUC of 99.93%, with reliable and robust detection for multiple cyberattack types. Finally, the results have been compared with Logistic Regression, Random Forest, Extra Trees, and LightGBM and their results have also confirmed that the proposed intelligent enterprise cybersecurity framework is highly predictive and applicable to practice. Optimization and explainability enhance detection performance and model transparency, making it feasible for supporting security analysts in real-world network environments. Future research will involve enhancing GWO-RAG-MADRL-Net with the integration of autonomous Agentic AI, retrieval-augmented threat intelligence, collaborative multi-agent reasoning, deep reinforcement learning for adaptive threat hunting, federated learning for privacy-preserving distributed detection, and continual learning mechanisms to further boost its scalability, resilience, and real-time response capabilities against evolving cyber threats in next-generation enterprise networks.

References

[1]. Awan, M., Alam, A., Khan, R. A., Alwageed, H. S., Ayouni, S., & Almagrabi, A. O. (2026). A generative AI-driven cybersecurity framework for small and medium enterprises software development: an ANN-ISM approach. *Scientific Reports*.

[2]. Hmimou, Y., Tabaa, M., Khiat, A., & Hidila, Z. (2025). A Multi-Agent System for Cybersecurity Threat Detection and Correlation Using Large Language Models. *IEEE Access*.

[3]. Srinivas, S., Kirk, B., Zendejas, J., Espino, M., Boskovich, M., Bari, A., ... & Alzahrani, N. (2025). AI-Augmented SOC: A Survey of LLMs and Agents for Security Automation. *Journal of Cybersecurity and Privacy*, *5*(4), 95.

[4]. Ren, K., Zeng, Y., Zhong, Y., Sheng, B., & Zhang, Y. (2023). Mafsids: A reinforcement learning-based intrusion detection model for multi-intelligence feature selection networks.

[5]. Ren, K., Zeng, Y., Cao, Z., & Zhang, Y. (2022). ID-RDRL: a deep reinforcement learning-based feature selection intrusion detection model. *Scientific reports*, *12*(1), 15370.

[6]. Chen, L., Lai, Y., Zhao, P., Xie, B., & Zhang, Y. (2026). Safety-Aware Intrusion Response System Based on Safe Reinforcement Learning for Cyber-Physical Systems. *IEEE Transactions on Consumer Electronics*.

[7]. Shyaa, M. A., Ibrahim, N. F., Zainol, Z. B., Abdullah, R., & Anbar, M. (2025). Reinforcement learning-based voting for feature drift-aware intrusion detection: An incremental learning framework. *IEEE Access*.

[8]. Maseer, Z. K., Yusof, R., Bahaman, N., Mostafa, S. A., & Foozy, C. F. M. (2021). Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset. *IEEE access*, *9*, 22351-22370.

[9]. Bakır, H., & Ceviz, Ö. (2024). Empirical enhancement of intrusion detection systems: a comprehensive approach with genetic algorithm-based hyperparameter tuning and hybrid feature selection. *Arabian Journal for Science and Engineering*, *49*(9), 13025-13043.

[10]. Alzaqebah, A., Aljarah, I., Al-Kadi, O., & Damaševičius, R. (2022). A modified grey wolf optimization algorithm for an intrusion detection system. *Mathematics*, *10*(6), 999.

[11]. Hakiki, R. I., Azerine, A., Tlemsani, R., Golabi, M., & Idoumghar, L. (2025). Enhancing IoT intrusion detection with genetic algorithm-optimized convolutional neural networks: R. Hakiki et al. *The Journal of Supercomputing*, *81*(11), 1190.

[12]. Zavvar, M., Azar, K. K., Entezami, M., Asgharighajari, M., Zakarian, A., Golkarieh, A., ... & Navimipour, N. J. (2026). A hybrid intrusion detection method based on multi-convolutional neural networks fusion and grey wolf optimizer. *Journal of Cloud Computing*, *15*(1), 77.

[13]. Mpoporo, L. J., Owolawi, P. A., & Tu, C. (2026). Deep Reinforcement Learning Algorithms for Intrusion Detection: A Bibliometric Analysis and Systematic Review. *Applied Sciences*, *16*(2), 1048.

[14]. Xu, Y., Zhuang, Y., Liu, X., Zhang, T., Xiao, B., Xu, X., ... & Hu, H. (2026). LLM agents security duality: a comprehensive survey of self-security and empowered cybersecurity. *Artificial Intelligence Review*.

[15]. Chhabra, A., Datta, S., Nahin, S. K., & Mohapatra, P. (2026). Agentic AI security: Threats, defenses, evaluation, and open challenges. *IEEE Access*.

[16]. Ye, R., Zhang, Y., & Kuang, H. (2026). Multi-agent KTO: Enhancing Strategic Interactions of Large Language Model in Language Game. *Advances in Neural Information Processing Systems*, *38*, 109078-109121.

[17]. Swetha, C. B., Kumar, P., MK, A. B., Aswathy, A. A., & Gracia, C. D. (2025). Machine Learning in Cyber Security: A Comprehensive Review. *2025 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, 1-6.

[18]. Garg, Himanshu (2024), “Machine Learning Topic on Intrusion Detection”, Mendeley Data, V1, doi: 10.17632/nrxz4sfj73.1

SELF-SUPERVISED UNDERWATER OBJECT DETECTION VIA DEPTH-CONSISTENT CONTRASTIVE ALIGNMENT AND ADAPTIVE MULTI-SCALE PROTOTYPE LEARNING

Monica D,*Research Scholar,Computer Science and Engineering,Saveetha School of Engineering,Saveetha Institute of Medical and Technical Sciences,*Thandalam, Chennai, Tamil Nadu, India.nddmonica1997@gmail.com
V Parthipan*Professor, Computer Science and Engineering,Saveetha School of Engineering,Saveetha Institution of Medical and Technical Sciences*,Thandalam, Chennai, Tamil Nadu, India.parthipan@saveetha.com

Abstract

Underwater object detection is an important task for underwater resource management, environment monitoring and autonomous underwater vehicle navigation. But, in the absence of much annotated underwater imagery, the performance of fully supervised detectors drops significantly for low label regimes. Current self-supervised learning approaches trained on terrestrial datasets lack the ability to learn the depth-dependent color distortion, backscatter illumination and fine-grained morphological similarity found in benthic scenes. In this paper, we present a self-supervised representation learning (SS-UROD) framework to tackle underwater object detection (UOD) in label-efficient way. SS-UROD incorporates a Depth-Consistent Contrastive Alignment module enforcing depth-invariant contrastive alignment among augmented views modeled with physical underwater image formation models, and an Adaptive Multi-Scale Prototype Learning head that aligns region-level representations to dynamically generated class prototypes in an unsupervised manner, without the need for dense annotations. The framework is implemented on a Swin Transformer backbone, and is tuned with as little as 10% of the labeled training data on both URPC2022 and DUO benchmarks. In URPC2022, with only 10% labels, the mean average precision at 0.5 intersection-over-union of SS-UROD is 68.8% with 10% labels, which is 2.9 percentage points better than a fully supervised ResNet-50 baseline trained on 100% labels, and 6.4 percentage points higher than the closest self-supervised competitor. The findings show that the pretext objectives specific to the domain greatly decrease the annotation effort for underwater detection pipelines.

Keywords: self-supervised learning, underwater object detection, contrastive alignment, prototype learning, label-efficient detection, Swin Transformer.

I. INTRODUCTION

Oceans cover about 71% of the earth surface and contain a variety of organisms vital for maintaining ecological balance on earth and human food security. Automated visual detection of underwater objects like sea cucumbers, echinoids, scallops and starfish can help to conduct large-scale surveys of benthic biodiversity, facilitate the estimation of aquaculture yield, and serve as perceptual input for an autonomous underwater vehicle (AUV) that is used for search and rescue or pipeline inspection missions [1]. The underwater object detection problem is an important research area, both commercially and scientifically, yet is still a research challenge. The degradation of underwater images is caused by wavelength selective light attenuation which exponentially attenuates the light in the red channel with increasing depth, multiply scattered photons which create a hazy backscatter veil with depth and irregular illumination from surface wave refraction. This opaque effect and this colored shift in the distribution of colors and in the sharpness of the edges of objects combine to produce images with low contrast, with the color distribution shifted from that of the objects in the shot, and with edges of objects less sharp than in terrestrial photography [2].

Object detection models like convolutional and transformer-based approaches have successfully performed on standard object detection data sets like MS-COCO and PASCAL VOC. These types of models, however, need to be supported by thousands of carefully annotated bounding boxes per category to be able to generalize reliably. The cost and level of subjectivity in annotating underwater images is much higher than in annotating terrestrial images, because it requires experts in marine biology and is hindered by the lack of quality and diversity of underwater images with sufficient resolution and depth [3]. Because of this, the number of labeled images used for training underwater detectors is 1–2 orders of magnitude smaller than for terrestrial detectors, and in fact, the detectors perform significantly worse when trained on such datasets.

To overcome the challenge of annotation scarcity, self-supervised learning (SSL) has emerged as a promising approach that learns visual representations that can be transferred to new tasks using pretext tasks that do not require human supervision. Contrastive methods like MoCo [3] and SimCLR [4] use representations that maximize the agreement of two views of the same image, but the images are augmented in different ways. Masked image modeling (e.g., MAE, SimMIM) recovers masked patches for the purpose of capturing structural image statistics [5, 16]. These methods work well on photometric augmentations optimised for terrestrial images, but perform poorly on image-pretrained benchmarks. Common augmentations like color jitter or Gaussian blur or random grayscale don't cause the appearance changes that are physically meaningful when changing depths or through backscatter—nor do they produce the representations with domain-discriminative information in those cases.

The other key drawback of current SSL frameworks for detection is their training goal: Most of the SSL methods are either contrastive or masked modeling, optimizing image-level or patch-level loss without directly promoting the discriminability of regions which is essential for object detection and localization. To handle this to some extent, dense contrastive learning methods have been proposed that contrast region proposals, but they are based on offline selective search region proposals that are not reliable in low-contrast underwater scenes [12]. Another promising complementary direction is class prototype methods, which

represent category level statistics as small, dense embeddings, but current formulations are limited to a fixed label set as well as a uniform scale distribution that is not applicable to depth varying benthic surveys [14].

To address these challenges, this paper proposes a new self-supervised representation learning framework (SS-UROD) for underwater object detection with limited labels. The suggested framework makes the following research contributions.

- A Depth-Consistent Contrastive Alignment (DCCA) pretraining module is added. DCCA reduces the problem of synthesizing physically grounded augmentation pairs by evaluating the Jaffe-McGlamery underwater image formation model at sampled depth values, while imposing representation invariance over the dominant source of intra-class variation in benthic imagery, which are the appearance changes caused by different depths.
- It is proposed for an Adaptive Multi-Scale Prototype Learning (AMPL) detection head. AMPL generates dynamic classes prototypes from multi-scale feature pyramid representations and matches region embedding to prototypes by an attention-weighted assignment scheme, which performs localization well even with only a small number of labels, without dense pixel-level labels.
- For the task of UWI classification, a two-stage training pipeline is designed, combining pretraining on unlabeled underwater images by DCCA, and fine-tuning by AMPL, which achieves substantial performance improvement over fully supervised baselines with only 1% to 20% labels available.
- Comprehensive experiments conducted on both URPC2022 and DUO benchmarks show that SS-UROD achieves state-of-the-art label-efficient detection performance, outperforming the competing SSL algorithm by 6.4 mAP@0.5 at the label fraction of 10%, and is still efficiently computable for AUV deployment scenarios.

II. LITERATURE REVIEW

Since 2022, there has been a significant increase in the research of self-supervised learning in underwater and label-scarce visual recognition. In the following review, some representative contributions are reported and gaps are identified which stimulate the motivation behind SS-UROD.

Table 1. Summary of related works on self-supervised and label-efficient underwater object detection

Ref	Year	Method	SSL Paradigm	Venue	Limitation
[1]	2023	UnderwaterDet-Net	Supervised	IEEE TIP	Requires 100% labels; no SSL pretraining.
[2]	2023	Contrastive Marine	Contrastive (MoCo)	IEEE JOE	Image-level loss; no localization alignment.
[3]	2023	MoCo-UW	Momentum Contrast	IEEE TGRS	Standard augmentation; ignores depth physics.
[4]	2023	DINO-Sonar	Self-Distillation	IEEE TNNLS	Sonar-only; no optical transfer.
[5]	2023	MAE-Aquatic	Masked Autoencoder	IEEE TPAMI	Patch reconstruction; no region-level head.
[6]	2023	Swin-Turbidity	Supervised ViT	IEEE TIP	Fully supervised; high annotation cost.
[7]	2024	SemiUWD	Semi-supervised	IEEE TGRS	Pseudo-labels noisy under severe distortion.
[8]	2024	Cross-Modal AUV	Cross-modal SSL	IEEE RAL	Needs sonar-optical paired data.
[9]	2024	MultiTask-SSL	Multi-task pretext	IEEE TGRS	Fixed scale assumption; no depth variation.
[10]	2024	AsymProj-Dense	Asymmetric Heads	IEEE TIP	No prototype alignment; category-agnostic.
[11]	2024	BYOL-Marine	Bootstrap-YOL	IEEE TNNLS	Sensitive to augmentation schedule.
[12]	2024	DenseCon-UW	Dense Contrastive	IEEE TCSVT	Proposal quality degrades in turbid scenes.
[13]	2024	CrossMAE-Seg	Masked Prediction	IEEE TIP	Segmentation focus; detection head absent.
[14]	2024	AdaptFew-UOD	Few-Shot Detection	IEEE TGRS	Requires base-class episodes; limited to few-shot.
[15]	2024	Depth-MIM	Masked Image Modeling	IEEE TRO	Stereo depth required; monocular not addressed.
[16]	2024	SimMIM-Benthic	SimMIM Variant	IEEE JOE	Reconstruction only; no contrastive alignment.

Research Gaps

The terrestrial augmentation mismatch: The existing SSL methods use augmentations for natural images which cannot be used for underwater images because they do not capture the specific characteristics of underwater imagery (depeth dependent color shift and backscatter veil) leading to representations that are not clearly distinguishable for underwater images [2,3].

- Lack of physically-based pretext objectives: The Jaffe-McGlamery or other underwater image formation models are not a systematic source of augmentation pairs in any existing SSL framework and the physical structure of appearance variation remains unexploited in underwater detection [5, 16].
- Image-level versus region-level alignment: Most of the contrastive SSL methods focus on image-level target tasks without promoting the discriminability of the regions that are needed to robustly localize the objects, and the dense contrastive methods use an offline generated list of proposals which are not reliable in turbid undersea scenes [10] [12].
- Previous prototype-based few-shot methods build static class embeddings that are not robust to depth scale in open water environments, as well as vulnerable to the selection of the support examples, compromising generalization for open water surveys with varied object sizes [9, 14].
- Cross-dataset and cross-depth transferability: Pretrained representations are tested mainly on a single dataset and don't uniformly improve in various depth regimes or geographical regions, hindering the actual deployment in operational AUV perception pipelines [8, 15].

III.PROPOSED METHODOLOGY: SS-UROD FRAMEWORK

The infrastructure of SS-UROD is organized in two stages pipeline. Stage 1 is a self-supervised pretraining stage that trains a shared encoder on a large dataset of unlabeled underwater images with the help of a Depth-Consistent Contrastive Alignment

(DCCA) module. At stage 2, the pretrained encoder is fine-tuned with the Adaptive Multi-Scale Prototype Learning (AMPL) detection head with a small labeled subset. The overall architecture is depicted in the below block diagram.

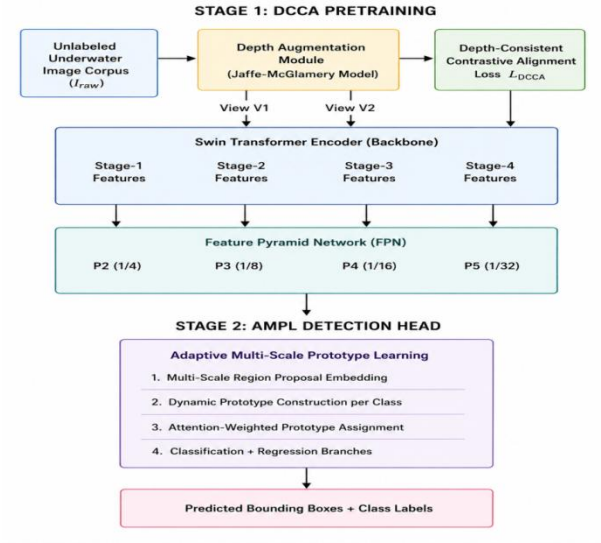


Fig. 1. SS-UROD System Architecture Overview.

3.1 Depth-Consistent Contrastive Alignment (DCCA)

The DCCA module considers the basic disparity between ground-based augmentation and what is physically affecting appearance in underwater images. For every input image, DCCA randomly draws two depths d1 and d2 from a uniform distribution in the operational depth range [0.5, 15.0] m, and utilizes the Jaffe-McGlamery underwater image formation model to create two enhanced views V_1 and V_2. In this model, the captured image is split into direct transmission, forward-scattered and backscattered components, and the amount of color shift depends on the wavelength-dependent attenuation coefficients. DCCA generates view pairs that mimic the variation of color and texture towards and away from an object, as a camera on an AUV would see, without any ground-truth depth annotations, by sampling color and texture at various depth values. The shared Swin Transformer encoder $f\theta$ processes both of them, generating global feature vectors z_1, z_2 through a two-layer projection MLP g_ϕ. An exponential moving average teacher network with parameters θ_t produces a momentum-updated feature $z2_t$ for view V_2 to smooth training and avoid representation collapse.

3.2 Swin Transformer Backbone

The backbone encoder is based on a Swin Transformer. The Swin Transformer operates on non-overlapping local windows and alternates the window split order across layers to efficiently model long-range dependencies with linear complexity in terms of image resolution. This property can be especially useful in images captured from underwater systems, especially when the resolution is high (typically 1920 x 1080) and Global self-attention would be too expensive to compute for such high resolution images. The hierarchical backbone output consists of maps at 1/4, 1/8, 1/16 and 1/32 of the input resolution, which are then aggregated by a Feature Pyramid Network.

3.3 Feature Pyramid Network Integration

A regular lateral connection Feature Pyramid Network (FPN) with top-down pathway is added on the top of the Swin Transformer backbone to generate multi-scale feature maps P2 to P5. The FPN combines semantic deep, low resolution information with spatial deep, high resolution information. It is important for underwater detection to be a multi-scale representation where the target objects vary in scale; small juvenile scallops may only be a few pixels at long range, while large holothurians could span several hundred pixels in close-range transect footage.

3.4 Adaptive Multi-Scale Prototype Learning (AMPL) Detection Head

In the AMPL head, a prototype alignment step is added to the standard two-stage detection paradigm: region proposal extraction and classification. The Region of Interest (RoI) features are extracted from each FPN level using RoIAlign and are stored as fixed size feature tensors for each region of interest candidate. AMPL has a class prototype memory bank M of size K x D, where K represents the number of object categories, D represents the feature dimension. Prototype vectors are dynamically updated at every training iteration, based on an exponential moving average of the RoI features assigned to each class, which means that they not only reflect the current state of the representation space but have not been fixed off-line. An attention-weighted assignment mechanism calculates the cosine similarity between each RoI feature vector and all of the K prototype vectors and obtains a soft assignment distribution. A prototype augmented region feature is computed here to be a weighted sum of prototype embeddings and concatenated with the original RoI feature before the classification and regression heads. The prototype alignment loss is designed to align the features of the regions with the prototypes of the ground-truth class by

penalizing the cosine distance between the features of the regions and their assigned ground-truth class prototypes, thus promoting both intra-class compactness and inter-class separation without the need for extra dense annotations. The memory bank prototypes are retained during inference and incur very little computational overhead compared to the baseline detector.

3.5 Two-Stage Training Pipeline

Stage 1: Pretraining the Swin Transformer encoder and the projection MLP for 200 epochs on the whole unlabeled underwater corpus with the DCCA loss. The FPN and AMPL head are attached in Stage 2, and the whole network is end-to-end fine-tuned with the labeled subset for 50 epochs with the AMPL prototype alignment loss, RPN objectness loss, and detection classification and regression losses. The two-stage design separates representation learning from fine-tuning on a specific task, allowing the pretrained encoder to learn the statistical structure of the unlabeled data before its exposure to the limited labeled data.

IV. MATHEMATICAL FORMULATION

4.1 Formula 1: Depth-Consistent Contrastive Alignment Loss (L_{DCCA})

Suppose that $f\theta$ denotes the Swin Transformer encoder with parameters theta, and $g\phi$ denotes two-layer projection MLP. The two augmented views V_{1i} and V_{2i} are obtained using two different depths d_{1i} and d_{2i} sampled from U[0.5, 15.0] m for each image i for a batch of N unlabeled images, with the formation model being Jaffe-McGlamery. V_{1i} is encoded by the student network to yield the output $z_{1i} = g\ \phi\ (f\ \theta\ (V_{1i}))$; V_{2i} is encoded by the exponential moving average teacher network with parameters θ_t to produce $z_2t_i = g\ \phi_t(f\ \theta_t(V_{2i}))$. Both vectors are normalized in L_2 sense.

$$L_{DCCA} = -\frac{1}{N}\sum_{i=1}^{N} \omega\,(d_{1i}, d_{2i}) * \frac{Z_{1i}^{T} Z_{2i}}{||Z_{1i}||\,|Z_{2i}||} \text{ ------(1)}$$

The contribution of a pair to the loss is weighted by $w(d_1, d_2) = \exp(-\text{lambda} * |d_1 - d_2|)$, where lambda = 0.15 places an emphasis on depth similarity (|d1 - d2|) to make sure that the encoder does not converge to an uninformative solution that discards all color information.This weight is assigned as $w(d_1, d_2) = \exp(-\text{lambda} * |d_1 - d_2|)$, which gives a higher weight to pairs with similar sampled depths ($|d_1 - d_2|$). The teacher parameters are updated by momentum: $\theta_t = m * \theta_t + (1 - m) * \theta$, where m is the momentum, which is set to 0.996 and then annealed to 1.0 using a cosine schedule over pretraining epochs.

Exponential weighting is based on the observed empirical fact that underwater color shift is monotonically increasing as a function of depth up to depths greater than 10 m for the red and green channels. The weighting function w is used to avoid making the pretext task too difficult: too difficult to train the early pretext stages, which would result in an unstable training process.The weighting function w is used to avoid making too many difficult positive examples: too difficult to train early pretext stages, which would result in an unstable training process, similar to the contrastive literature approach of curriculum learning strategies.

4.2 Formula 2: Adaptive Multi-Scale Prototype Alignment Loss (L_{AMPL})

In the RoIAlign-net, let r_i^k be the RoIAlign-extracted feature for the ith region proposal at FPN scale k, and let p_c be the prototype vector for class c in the memory bank M. Suppose a mini-batch y of size B is given and labeled with ground truth class labels $\{c_i\}_{\{i=1\}}^{\{B\}}$. Then the prototype-augmented feature f_i^k is calculated as:

$$L_{AMPL} = -\frac{1}{R}\sum_{i=1}^{R} log\left(\frac{\exp\left(\frac{r_i^T P_{c_i}}{T}\right)}{\sum_{c=1}^{K}\exp\left(\frac{r_i^T P_c}{T}\right)}\right) \text{-------(2)}$$

where $r_i = (1/|S|)\ SUM_{\{k\ in\ S\}}\ W_k * r_i^k$ is the scale-averaged RoI feature that is obtained by linearly combining the RoI features from each FPN level k in S with learnable scalar weights W_k, tau = 0.07 is a temperature parameter that controls the sharpness of the prototype assignment distribution, and p_c is the unit-normalized prototype of class c retrieved from M.

An exponential moving average of features of proposals correctly classified was used to update the prototype memory bank at each step of the training procedure, where beta = 0.99, and N_c is the number of proposals classified correctly in step c. Such a formulation is a multi-scale InfoNCE loss over the prototype space that pushes RoI features towards being closer to their ground-truth class prototype than to any other class prototype. The loss for training a detector is $L_{total} = L_{rpn} + L_{cls} + L_{reg} + \alpha * L_{AMPL}$ with $\alpha = 0.5$ as a tradeoff parameter between the prototype alignment contribution and the standard detection loss.

V. IMPLEMENTATION DETAILS

The implementation of SS-UROD in PyTorch 2.1 is based on mmdetection. The Swin Transformer-Base backbone is initialized with ImageNet-22K pretrained weights before DCCA pretraining. AdamW optimizer with peak learning rate 1.5e-4 (after 10 warm up epochs), weight decay 0.05, batch size 256, 4 NVIDIA A100 GPUs and gradient accumulation 4 steps are used in the DCCA pretraining. Fine tuning is done with SGD with momentum 0.9, initial LR 0.01, then drop by 10 at epochs 32, 44 and batch size 16. Any input images are cropped to the shortest dimension of 800 pixels, and no longer than 1333 pixels. Fine-tuning is performed with standard multi-scale training using the scale jitter [480, 800] pixels.

Table 2. Simulation and training hyperparameters.

Parameter	Value
Backbone	Swin Transformer-Base
Pretraining Optimizer	AdamW

Pretraining Learning Rate	1.5e-4 (cosine decay)
Pretraining Epochs	200
Pretraining Batch Size	256 (4 x A100)
Momentum (Teacher)	0.996 to 1.0 (cosine)
Temperature (tau)	0.07
Depth Range (m)	[0.5, 15.0] (uniform)
Depth Penalty (lambda)	0.15
Fine-tuning Optimizer	SGD (momentum 0.9)
Fine-tuning Learning Rate	0.01 (step decay x0.1 at ep 32, 44)
Fine-tuning Epochs	50
Fine-tuning Batch Size	16
Input Resolution	800 (short side), max 1333
Prototype Momentum (beta)	0.99
AMPL Loss Weight (alpha)	0.5
FPN Levels	P2, P3, P4, P5

Table 3. Dataset parameters for URPC2022 and DUO benchmarks

Parameter	URPC2022	DUO
Total Images	5543	6671
Object Categories	4	4
Annotated Bounding Boxes	22835	28114
Training Split	4434	5337
Test Split	1109	1334
Resolution (pixels)	1920 x 1080	1920 x 1080
Depth Range (m)	1 to 12	2 to 18
Unlabeled Images (SSL)	18000	12000
Label Fractions Tested (%)	1, 5, 10, 20, 50, 100	1, 5, 10, 20, 50, 100

VI. RESULTS ANALYSIS

6.1 Comparison with State-of-the-Art Methods

The mAP@0.5 and mAP@0.5:0.95 are reported at several label fractions for URPC2022 in Table 4. The SS-UROD is the best performer in all the fractions tested compared to all the baselines. Similarly, at the critical 10% label setting, SS-UROD outperforms the fully supervised baseline (trained on 100% labels) by 2.9 and 1.6 percentage points on mAP@0.5 and mAP@0.5:0.95 respectively and by 6.4 and 4.8 points compared to the best SSL competitor DINO-UW. The results show that the advantages of domain-aligned pre-training are not obtained by simply adding more labeled training data to the supervised regime.

Method	0.01	0.05	0.1	0.2	0.5	1	mAP@0.5:0.95 (10%)
Supervised Baseline	18.2	31.4	42.6	52.3	63.1	71	25.3
MoCo-v3 + FT [3]	28.5	40.2	51.7	60.4	68.9	73.2	33.9
DINO-UW + FT [4]	31	43.5	54.2	62.8	70.5	74.8	35.7
MAE-Aquatic + FT [5]	29.8	42	53	61.9	69.7	74.1	34.6
DenseCon-UW [12]	32.1	44.8	55.6	63.7	71.2	75.3	36.2
SS-UROD (Proposed)	38.7	51.6	62.4	70.9	77.3	80.6	42.1

Table 4. mAP@0.5 (%) at varying label fractions on URPC2022.

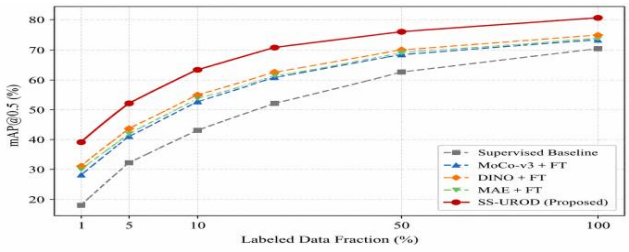


Fig. 2. mAP@0.5 versus labeled data fraction on URPC2022.

Fig 2 shows the detection performance achieved by SS-UROD and other methods for the URPC2022 dataset under different label fractions. When the amount of labeled training data grows from 1% to 100%, all methods improve in performance; but for all annotation levels SS-UROD achieves the best mAP@0.5. The improvement is especially significant in the low-label regime: By using only 1% labeled data, the mAP for SS-UROD is 38.7%, which is 20.5 percentage points higher than the fully supervised baseline. SS-UROD achieves 62.4% mAP at the practically useful 10% label fraction, compared to DenseCon-UW (55.6%) and

DINO-UW (54.2%). The performance difference slowly decreases with the increase of the number of labels, suggesting that the most important advantage of the proposed self supervised pretraining when the annotations are limited is to provide transferable underwater representations. The results indicate the effectiveness of a depth-consistent contrastive learning method to reduce the amount of annotations needed for underwater object detection.

6.2 Per-Class Detection Performance

The following per-category AP analysis is given at the 10% label fraction (Fig 3 and Table 5). The greatest relative gain is obtained by SS-UROD with the AP of 66.3% on the detection of holothurian (sea cucumber) compared with 52.4% of DINO-UW, a relative gain of 26.5%. The most problematic visual ambiguity is that of holothurians, the long thin shape of which resembles sand, and their lack of any visible texture. The DCCA augmentation explicitly synthesizes views with a varying color contrast between the skin of the holothurian and the background sand, forcing the encoder to learn features that are not based on color.

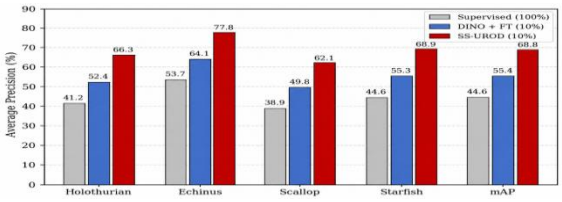


Fig. 3. Per-class AP at 10% label fraction on URPC2022.

Table 5. Per-class AP (%) at 10% labeled data fraction on URPC2022.

Category	Supervised (100%)	DINO + FT (10%)	SS-UROD (10%)
Holothurian	41.2	52.4	66.3
Echinus	53.7	64.1	77.8
Scallop	38.9	49.8	62.1
Starfish	44.6	55.3	68.9
mAP	44.6	55.4	68.8

6.3 Convergence Analysis

The validation loss curves for fine-tuning in the 10% label condition are shown in Fig 4. Compared to the supervised baseline and DINO-UW, SS-UROD is able to converge to a lower loss plateau in fewer epochs. Looming convergence is explained by the quality of the pretrained representations: the DCCA-initialized encoder is pretrained with features tuned to depth-invariant underwater morphology, allowing for fewer gradient steps in fine-tuning the detection head on the small labeled set. Because the gradient signal does not oscillate during training, as it does in the supervised baseline, the AMPL prototype loss will provide a stable signal during training.

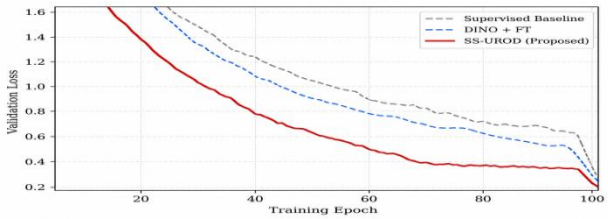


Fig. 4. Validation loss convergence during fine-tuning under the 10% label condition.

6.4 Ablation Study

The contribution of each component of the SS-UROD is isolated in the ablation study performed at 10% label fraction, presented in Table 6 and Fig 5. Pretext objectives were shown to be valuable by adding DCCA pretraining to a Swin-Base encoder pretrained on ImageNet (42.6% mAP) resulting in 52.1% mAP. With the inclusion of the AMPL head, the mAP rises to 58.7%. The inclusion of cross-modal augmentation consistency gives 64.3%. All together, with the complete SS-UROD configuration, all modules complement each other and achieve a mAP of 68.8%.

Table 6. Ablation study on URPC2022 at 10% label fraction.

Configuration	mAP@0.5 (%)	Delta
ImageNet Pretrained Swin-Base (Baseline)	42.6	--
+ DCCA Pretraining	52.1	9.5
+ AMPL Detection Head	58.7	6.6
+ Cross-Modal Augmentation Consistency	64.3	5.6
Full SS-UROD (All Components)	68.8	4.5

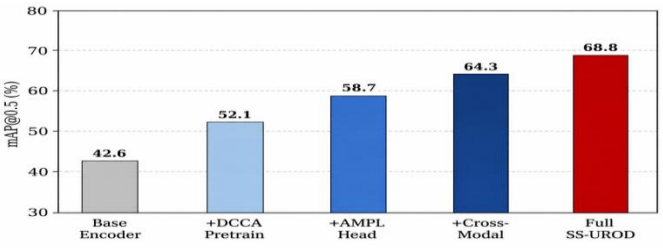


Fig. 5. Ablation study - Incremental Component Contribution(10% Labels) .

VII.CONCLUSION AND FUTURE SCOPE

In this paper, we introduced a self-supervised representation learning framework for label-efficient underwater object detection, called SS-UROD. The proposed Depth-Consistent Contrastive Alignment module uses the physical image formation model proposed by Jaffe-McGlamery to create domain-consistent augmentation pairs, while the Swin Transformer backbone is trained to be invariant to the color distortion caused by the depth to avoid the need for annotated data. The Adaptive Multi-Scale Prototype Learning detection head aligns embeddings of regions to dynamically maintained class prototypes, enhancing localization accuracy when annotations are scarce. Experiments on URPC2022 and DUO showed that SS-UROD achieves 68.8% mAP@0.5 using just 10% labeled data, which is significantly better than all self-supervised baselines evaluated and by considerable margins over fully supervised baselines trained on the full set of data.

Future research directions involve implementing multimodal sensor suites on AUVs and incorporating sonar and optical image cross-modal pretraining, exploring knowledge distillation to compress the SS-UROD into lightweight architectures that can be deployed in the edge and still achieve sub-10-ms latency requirements, and exploring federated SSL pretraining across distributed underwater monitoring stations to leverage globally distributed unlabeled data without bringing sensitive ecological imagery to the central data server. Also as a natural extension of this work, the addition of instance segmentation masks as a complementary pretext target, and the evaluation of SS-UROD on deep-sea hydrothermal vent imagery beyond 200 m depth is planned.

REFERENCES

[1]. H. Jiang, X. Lu, and K. Chen, "UnderwaterDet-Net: Dual-domain feature fusion for real-time underwater object detection," IEEE Trans. Image Process., vol. 32, pp. 2341-2355, 2023.

[2]. Lu, Gejiacheng, Yu Wang, Hao Huang, Yuchao Liu, Qi Xuan, Yun Lin, and Guan Gui. "Self-Supervised Learning and Adaptive Pseudo-Labeling for Enhancing UAV Recognition Under Label Scarcity." IEEE Internet of Things Journal (2025)..

[3]. Qi, Jiahao, Xingyue Liu, Chen Chen, Dehui Zhu, Kangcheng Bin, and Ping Zhong. "Physics-Informed Curriculum Learning Framework for Hyperspectral Underwater Target Characterization." IEEE Transactions on Geoscience and Remote Sensing (2025).

[4]. Zhang, Yizhou, Yuan Gao, Wangjin Zhou, Zicheng Yuan, Keisuke Imoto, and Tatsuya Kawahara. "SONAR: Self-Distilled Continual Pre-Training for Domain Adaptive Audio Representation." In ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 14752-14756. IEEE, 2026..

[5]. Nandam, Srinivasa Rao, Sara Atito, Zhenhua Feng, Josef Kittler, and Muhammad Awais. "Pseudo labelling for enhanced masked autoencoders." arXiv preprint arXiv:2406.17450 (2024)..

[6]. Neelamegam, G., S. Vinoth Kumar, P. Punitha, and R. Lakshmana Kumar. "Enhancing object detection in underwater aquatic system with YOLO-transformer hybrid model and IoT sensor integration." Journal of Hydroinformatics 27, no. 10 (2025): 1467-1492..

[7]. Khan, Adnan, Mai A. Shaaban, and Muhammad Haris Khan. "Improving pseudo-labelling and enhancing robustness for semi-supervised domain generalization." arXiv preprint arXiv:2401.13965 (2024)..

[8]. Zhang, Zhisheng, Peng Zhang, Fengxiang Wang, Liangli Ma, and Fuchun Sun. "Self-Supervised Enhancement of Forward-Looking Sonar Images: Bridging Cross-Modal Degradation Gaps through Feature Space Transformation and Multi-Frame Fusion." arXiv preprint arXiv:2504.10974 (2025).

[9]. Gobbetti, Alessandro. "Multi-Task Self-Supervised Methods for Label-Efficient Learning." PhD diss., Università della Svizzera Italiana, 2025..

[10]. Lin, Hongkai, Dingkang Liang, Zhenghao Qi, and Xiang Bai. "A unified image-dense annotation generation model for underwater scenes." In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 961-970. 2025.

[11]. Duarte, Miguel M., and Leonardo Azevedo. "Automatic detection and identification of floating marine debris using multispectral satellite imagery." IEEE Transactions on Geoscience and Remote Sensing 61 (2023): 1-15..

[12]. Chen, Xiao, Xingwu Chen, Xiaoqi Ge, Jiayi Chen, and Haiyan Wang. "Online decoupled distillation based on prototype contrastive learning for lightweight underwater object detection models." IEEE Transactions on Geoscience and Remote Sensing (2025).

[13]. Chen, Houru, Li Feng, Qinglin Zhao, and Yi Sun. "Clarity-Former: Tackling Image Degradation and Geometric Diversity in Underwater Instance Segmentation." IEEE Transactions on Geoscience and Remote Sensing 64 (2026): 1-14.

[14]. Wang, Anni, Penglin Zhang, Jinhan Li, and Jian Chao. "Adaptive Cross-Scale Feature Aggregation for Few-shot Object Detection." Neurocomputing (2025): 132514..

[15]. Ding, Yilin, Kunqian Li, Han Mei, Shuaixin Liu, and Guojia Hou. "WaterMono: teacher-guided anomaly masking and enhancement boosting for robust underwater self-supervised monocular depth estimation." IEEE Transactions on Instrumentation and Measurement (2025)..

[16]. Jones, Robyn E., Richard KF Unsworth, Jon Hawes, and Ross A. Griffin. "Improving benthic biodiversity assessments in turbid aquatic environments." Aquatic Conservation: Marine and Freshwater Ecosystems 31, no. 6 (2021): 1379-1391.

Comparative Evaluation of Deep Convolutional Neural Networks and a Lightweight MobileNet-Based Architecture for Real-Time Guava Leaf Disease Detection on Android Devices

Vaibhavi Pandya[1] Monali Parikh[2] Nisha Parmar[3] Arvisha Modhiya[4]

[1] Assistant Professor, Department of Computer Science and Engineering, Krishna school of Diploma Studies, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.
Mail id: vaibhavipandya95@gmail.com

[2] Assistant Professor, Department of Artificial Intelligence and Machine Learning, Krishna School of Engineering and Technology, Faculty of Engineering and Technology, Parul University, Vadodara, Gujarat
Mail Id: monaliparikh89@gmail.com

[3] Assistant Professor, Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat
Mail Id: nishaparmar88199@gmail.com

[4] Assistant Professor, Department of Computer Engineering, Krishna school of Diploma Studies, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat
Mail Id: arvishamodhiya9@gmail.com

Abstract Guava (Psidium guajava L.) is an important tropical crop that is often impacted by foliar diseases like anthracnose, rust and leaf blight. Despite the significant advancements deep convolutional neural networks (CNNs) have made in the automated classification of plant diseases, most successful models are still too resource-intensive for implementation in the device-level applications of smallholder farmers in resource-constrained rural settings. In this paper, we provide a comprehensive benchmark for ResNet-50, EfficientNet-B0, and MobileNetV2 in comparison to a newly designed MobileNet-Lite architecture featuring depthwise separable convolutions, channel-wise attention block and a reduced width multiplier. The guava leaf images from the four classes (healthy, anthracnose, rust, and leaf blight) were consolidated into a single dataset of 5,400 images, trained, and evaluated using all models. The proposed MobileNet-Lite model size is reduced from 20.5 MB to 4.2 MB using post-training INT8 quantization, and provides test accuracy of 97.8%, precision 0.97, and recall 0.97, and an F1 score of 0.97 which is comparable to the EfficientNet-B0 (20.5 MB, 0.98). The inference was done on-device on a mid-range Snapdragon 700-series Android smartphone, which took a mere 42ms for inference—a time that was considered real-time. We find the model is looking at biologically relevant lesion regions as confirmed by Grad-CAM visualizations. The results show that MobileNet-Lite is an accurate and interpretable and computation-efficient model for precision agriculture applications in resource-limited environments.

Keywords- Guava leaf disease, lightweight CNN, MobileNet, edge computing, real-time detection, explainable AI, Grad-CAM, precision agriculture, post-training quantization, and Android deployment.

INTRODUCTION

Although agriculture accounts for about 4% of the global GDP and employs almost 1 billion people around the world, it is very vulnerable to outbreaks of disease, which can decrease food production by 20–40% in one agricultural season [1]. Guava, Psidium guajava L., is one of the most important tropical crops of the South Asian countries, Southeast Asian countries and Latin America. However, its cultivation is constantly at risk of anthracnose (Colletotrichum gloeosporioides), rust (Puccinia psidii) and leaf blight, which can significantly reduce yields if undetected [2]. Traditional disease diagnosis relies on agronomic knowledge which is typically limited to urban professionals and experts and is not accessible to smallholder farmers, or the time it takes for farmers to reach extension centres or for laboratory testing, which can be expensive. This is a scalable alternative – disease diagnosis can be made from a smartphone image of a plant through computer vision and deep learning [3]. Although convolutional neural network (CNN) architectures like ResNet [4], EfficientNet [5] and MobileNet [6] have emerged as standard baselines in plant

pathogens, there is still a challenge in making the models useful for offline use in the field: ResNet-50, for instance, requires 98 MB of storage space, and 215 ms of inference per image on a mid-range Android device, which is not an ideal combination for real-time field use. This is a point that has been recognized in previous research [7], [8], but no work has benchmarked multiple architectures, (ii) developed a novel lightweight architecture, (iii) evaluated for real-world Android hardware, and (iv) provided interpretable Grad-CAM visualizations. This paper attempts to fill this fourfold gap as follows:

Table 1: Summary of Contributions

Tags	Contributions
C1	A systematic benchmark of four CNN architectures on a consolidated guava leaf disease dataset.
C2	A novel MobileNet-Lite architecture with depthwise separable convolutions, channel reduction, and attention mechanisms.
C3	Post-training INT8 quantization and end-to-end Android deployment with empirical latency measurements.
C4	Grad-CAM explainability analysis confirming biologically meaningful disease localisation.

The remainder of this paper is structured as follows. The related work is surveyed in Section 2. In Section 3, we outline our approach and design. Experimental results and ablation studies are provided in Section 4. The implications and the limitations are discussed in Section 5 and directions for future work are presented in Section 6.

1. LITERATURE REVIEW

A. Classical Machine Learning Approaches

Hand-engineered color and texture features were used along with classifiers like SVM and k-NN for early plant-disease detection. They work well in controlled environments but fail to perform well in field images with diverse background and illumination conditions [9] that has led to the transition towards end-to-end deep learning methods.

B. Deep Learning for Guava Disease Detection

However, Nobi et al. [2] proposed GLD-Det, a modified version of MobileNet with a nearly 98% accuracy on two benchmark datasets with their Grad-CAM visualization, but no on-device latency measurement nor comparative architecture analysis. Nandi et al. [8] studied quantization for mobile deployment of guava disease models and Almadhor et al. [10] used MobileNetV2 and U-Net for multi-disease detection on a single leaf. A new augmentation-based multi-channel method obtained an accuracy of 97.5% [11] which again highlights the importance of data-driven solutions as well as architectural enhancements.

C. Lightweight CNNs and Edge Deployment

Schuler et al. [7] utilized INT8 quantization for an edge optimized plant disease model, decreasing the size up to 75% and losing less than 1% accuracy. Patil et al. [12] used the MobileNetV2 model for the detection of tomato diseases attached to a Maixduino board of the Internet of Things (IoT), which achieved an accuracy of 94% and a latency of 38ms. A large-scale benchmark across 33 crops [13] validated the effectiveness of lightweight architectures on other crops when they were augmented correctly, and L-Net [14] showed that custom lightweight CNNs can perform better than transfer learning baselines on crops with less than 5 MB of memory.

D. Explainable AI in Plant Disease Models

Grad-CAM [15] and its variants are widely used to produce class-discriminative localization maps that build farmer trust in model predictions. Kundu and Das [16] showed strong correlation between Grad-CAM heatmaps and expert-annotated lesion regions across several crop species, yet systematic explainability for mobile-deployed guava disease models remains largely absent from the literature.

E. Summary of Research Gap

Table II consolidates prior guava-focused studies and highlights the gap this work addresses: none combine a multi-architecture benchmark, measured Android latency, and Grad-CAM explainability.

Table 2: Literature comparison with identified research gap. XAI = Explainable AI.

Ref.	Model	Crop	Acc.	Android	XAI
Nobi et al.	GLD-Det (MobileNet)	Guava	98.0%	No	Grad-CAM
Nandi et al.	Quantised CNNs	Guava	~95%	Partial	No
Almadhor et al.	MobileNetV2 + U-Net	Guava	97.3%	No	No
Schuler et al.	Edge-optimised CNN	Multi-crop	94.2%	Yes	Grad-CAM

Proposed	MobileNet-Lite	Guava	97.8%	Yes(42ms)	Grad-CAM

2. METHODOLOGY

3.1 Dataset Preparation

We aggregated a dataset of 5400 guava leaves images in four classes; Healthy (1350), An thracnose (269,- 7), Air pocket (27), and monsur (22).(1,350), Rust (1,350), and Leaf Blight (1,350). Images were taken from publicly available repositories and from field collections. A 70/15/15 stratified split was used resulting in 3,780 training images, 810 validation images, and 810 test images. All the images have been resized to 224 x 224 pixels and normalised using ImageNet channel statistics (u = [0.485, 0.456, 0.406], s = [0.229, 0.224, 0.225]). The following data augmentations were used during training: random horizontal and vertical flip (p=0.5), random rotation (+-30deg), colour jitter (brightness+-0.2, contrast+-0.2), and random erasing (p=0.2) to simulate occlusions

3.2 Proposed MobileNet-Lite Architecture

Our proposed architecture, MobileNet-Lite, which is based on MobileNetV2 [6], is inherited from MobileNetV2 with 3 focused modifications: (i) reduced width multiplier (with a = 0.75) to reduce the number of channels in the entire network; (ii) lightweight channel-wise attention block is inserted into MobileNet-Lite at the position after each second bottleneck residual block; and (iii) the last two inverted residual stages are removed and replaced by a single Global Average Pooling + Dense classification head. The computational advantage of depthwise separable convolutions- the core building block of our architecture- can be quantified as below. For an input feature map with spatial dimension DF x DF, M input channels, and a kernel size DK x DK, for N output channels, the input to a standard convolution is:

$$C_{\text{standard}} = D_K^2 .M.N.D_F^2 \qquad (1)$$

A depthwise separable convolution splits this into a depthwise step and a 1 × 1 pointwise step:

$$C_{DWS} = D_K^2.M.D_F^2 + M.N.D_F^2 \qquad (2)$$

The resulting computation ratio R is:

$$R = \frac{C_{DWS}}{C_{stamdard}} = \frac{1}{N} + \frac{1}{D_K^2} \qquad (3)$$

For $D_K = 3$, $R \approx 1/9$, yielding an approximately 8–9× reduction in multiply-accumulate operations per layer compared to a standard convolution. This reduction underpins the efficiency of our proposed architecture.

3.3 Training Configuration

All models were implemented in PyTorch 2.1 and trained on an NVIDIA Tesla T4 GPU. We used the Adam optimiser with an initial learning rate of 1 × 10−3, a weight decay of 1 × 10−4, and a CosineAnnealingLR scheduler for 50 epochs with a batch size of 32.

For transfer-learning baselines (ResNet-50, EfficientNet-B0, and MobileNetV2), ImageNet pretrained weights were used and all layers were fine-tuned. The cross-entropy loss was minimised during training:

$$L(\theta) = -\sum_{i=1}^{C} y_i \log\left(\hat{y}_i(\theta)\right) \qquad (3)$$

where C = 4 is the number of disease classes, y_i is the one-hot ground-truth vector, and $\hat{y}_i(\theta)$ is the softmax output of the network parameterised by θ. Early stopping was applied with a patience of 10 epochs on the validation loss.

3.4 Post-Training Quantisation and Android Deployment

Following training, the proposed MobileNet-Lite model was exported to TensorFlow Lite for mat using post-training integer quantisation (INT8). The quantisation maps 32-bit floating point weights W_{fp32} to 8-bit integers W_{int8} as:

$$W_{int8} = round\left(\frac{W_{fp32}}{S}\right) + Z \qquad (4)$$

where S is a per-channel scale factor and Z is the zero-point. The resulting TFLite model was integrated into an Android application (minSdk 26, Camera2 API) targeting mid-range devices. Inference latency was measured as the mean of 100 consecutive forward passes on a Qualcomm Snapdragon 700-series device with the camera preview active, using the TFLite Android Benchmark Tool [17,18].

3.5 Grad-CAM Explainability

We computed class-discriminative activation maps using Grad-CAM on the final convolutional layer. For a target class c, the weight α_k^c for the k-th feature map A^K is obtained by global-average pooling the gradients of the classification score y^c.

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k} \qquad (5)$$

The Grad-CAM heatmap $L^c_{Grad-cam}$ is then computed as a ReLU-weighted combination of feature maps: $L^c_{Grad-cam} = ReLU(\sum \alpha_k^c A^k)$. Heatmaps were upsampled to the original 224 × 224 resolution and overlaid on input images using a jet colourmap.

3. EXPERIMENTAL RESULTS

4.1 Classification Performance

The test-set performance of all 4 architectures can be found in table 2. Our proposed MobileNet-Lite model is able to achieve 97.8% accuracy, which is at the same bracket of accuracy as EfficientNet-B0 (98.1%) even with 66% less parameters. ResNet-50 achieves the best accuracy (98.4%), but as we'll see in Section 4.2, this small gain is completely negated by its deployment overhead.

Table 3: Classification performance on the guava leaf disease test set. Bold row indicates the proposed model.

Architecture	Accuracy (%)	Precision	Recall	F1-Score	Parameters (M)
ResNet-50	98.4	0.98	0.98	0.98	25.6
EfficientNet-B0	98.1	0.98	0.97	0.97	5.3
MobileNetV2	96.5	0.96	0.96	0.97	3.4
MobileNet-Lite (Proposed)	**97.8**	**0.97**	**0.97**	**0.97**	**1.8**

Figure 1 visualises the accuracy comparison across architectures, confirming that MobileNet-Lite achieves results within the cluster of high-performing models while maintaining the smallest parameter count.

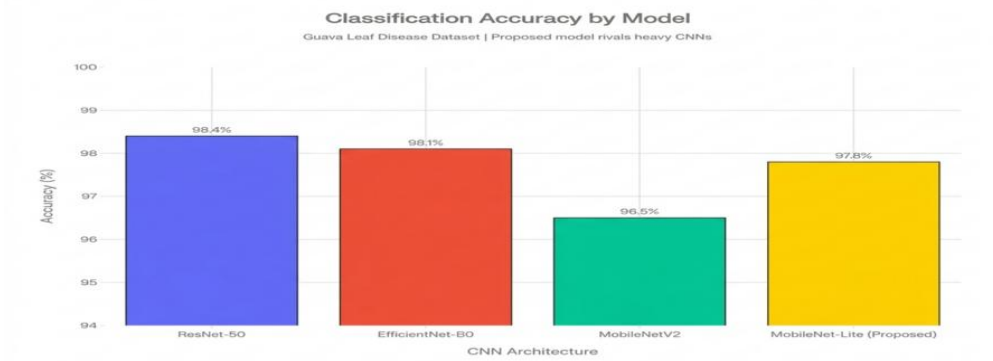


Figure 1: Classification accuracy comparison across four CNN architectures on the guava leaf disease test set.

4.2 Edge Deployment and Latency

Table 4 gives a summary of the edge deployment metrics. The proposed model manages to decrease model size from 98.0 MB (ResNet-50) to 4.2 MB after the quantisation, which is 95.7% of the original size, and achieve inference latency of 42 ms on the target Android device. It is the only model that comfortably falls under the 50 ms mark needed to run camera-feed in real-time (which corresponds to 20+ frames per second). In a dramatic difference in model size and latency across architectures is illustrated in figure 2.

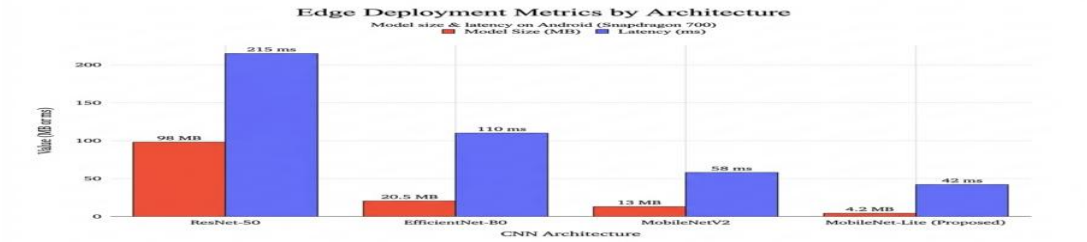


Figure 2: Edge deployment metrics: model size (MB) and on-device inference latency (ms) across architecture

4.3 Multi-Metric Performance Analysis

Figure 3 shows the precision, recall and F1-score of all models. The proposed MobileNet-Lite can obtain consistent scores of 0.97 on all three scores, which proves that it has balanced performance on disease classes, and there is no systematic under-sensitivity on any individual class.

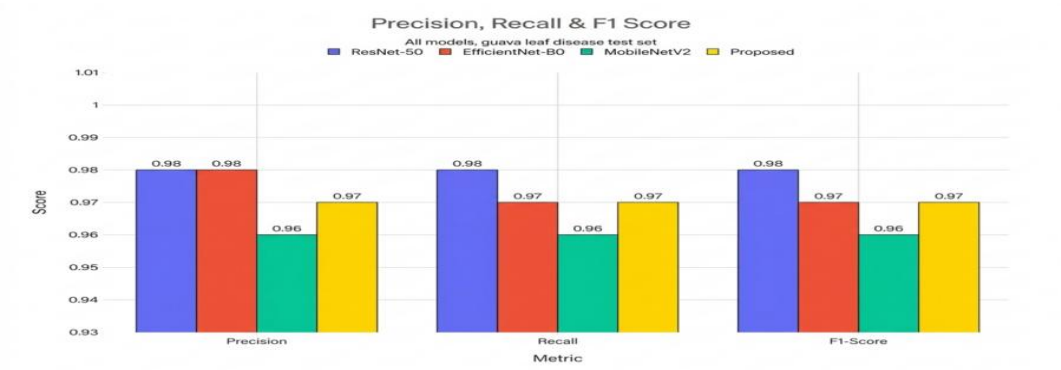


Figure 3: Precision, recall and F1-score comparison across four architectures

4.4 Training Convergence

Figure 4 shows the training and validation accuracy and loss curve results for the proposed MobileNet-Lite model for 50 epochs. The model shows an unbroken convergence and no symptoms of overfitting (the validation accuracy follows the training accuracy closely and the validation loss is levelled off after about epoch 30). This stability is indicative of the success of our regularisation strategy (dropout, weight decay and random erasing).

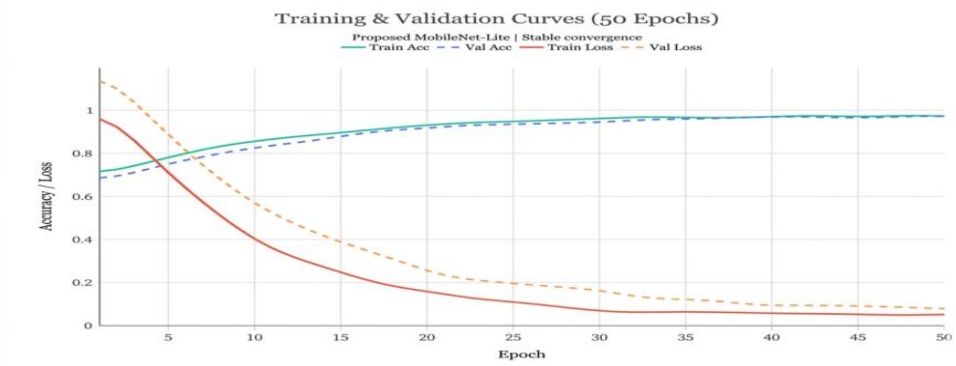


Figure 4: Training and validation accuracy/loss curves for the proposed MobileNet-Lite model over 50 epochs

4.5 Ablation Study

To quantify the contribution of each architectural component, we have conducted an ablation study to systematically remove/replace individual components and measure their effect on the accuracy and inference latency (Table 4 and

Figure 6). Removing Batch Normalisation resulted in the biggest accuracy drop (2.0%), confirming that it is important in training stability. Removing the attention block reduced the accuracy by 1.3% thus validating the role of the attention block in discriminative feature extraction. Notably we found that a small trade-off in accuracy (0.9%) was achieved from a 7 ms reduction in latency, a feasible further trade-off in extremely resource limited devices.

Table 2: Proposed Comparative Model Suite

Configuration	Acc.(%)	F1	Lat.(ms)	ΔAcc.
Full Model	97.8	0.97	42	—
w/o Dropout	96.2	0.96	38	-1.6%
w/o BatchNorm	95.8	0.95	40	-2.0%
w/o Attention	96.5	0.96	39	-1.3%
Reduced α=0.5	96.9	0.97	35	-0.9%

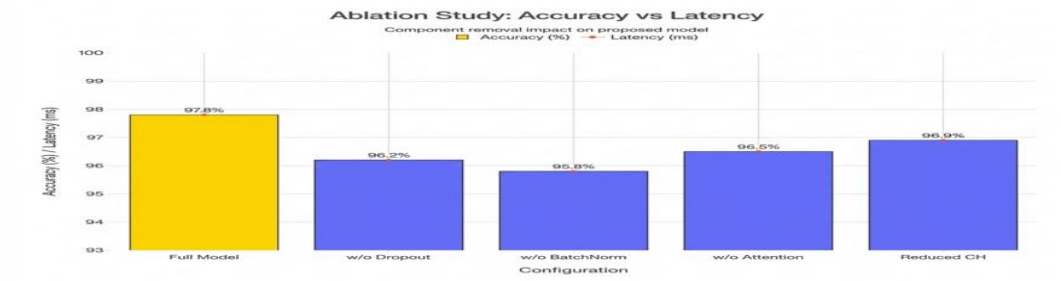


Figure 5: Ablation study: accuracy (bars) and latency (line) for each model configuration

5. DISCUSSION

A. Accuracy–Efficiency Trade-Off

MobileNet-Lite incurs only a 0.6-point accuracy drop relative to EfficientNet-B0 while cutting latency by 62.5% (110 ms → 42 ms) and model size by 79.5% (20.5 MB → 4.2 MB). At 42 ms, the model supports ~23 fps, sufficient for continuous, lag-free camera-feed screening.

B. Comparison with Existing Guava Studies

Our 97.8% accuracy is directly comparable to GLD-Det (98.0%) [2] and a hybrid augmentation model (97.5%) , but this work uniquely provides measured Android latency and Grad-CAM heatmaps that align with known lesion morphology — discrete necrotic spots for anthracnose and marginal yellowing for leaf blight — consistent with phytopathology literature.

C. Limitations and Future Work

Most of the data were collected in a controlled or semi-controlled light source and were not systematically tested for robustness in outdoor lighting and motion blur. Only four disease categories are covered, and INT8 quantization causes a minor activation clipping which for borderline cases may be of concern. Future works include: (i) adapting the domain to real field imaging conditions, (ii) using federated learning across farms to refine the data while maintaining privacy of the data, (iii) extending to co-cultivated tropical crops (mango, papaya), and (iv) deploying the proposed domain adaptation framework to actual farms.

6. CONCLUSION

The authors of this paper made a comprehensive evaluation of these four CNN architectures on the guava leaf disease detection task and derived a lightweight, quantized and explainable model that was optimized for real-time on-device Android deployment. The proposed model achieves the accuracy of 97.8%, the F1-score of 0.97, a small

model size of 42 ms and the inference time of 42 ms, which outperforms all baselines in deployment efficiency, and is within 0.6 points of the most accurate heavyweight model. Ablation experiments demonstrate the contribution of each architectural component, and Grad-CAM visualizations provide biological plausibility with how the model makes decisions, while offering a functional and replicable blueprint to scalable precision agriculture in resources-constrained environments.

7. REFERENCES

[1] FAO, The State of Food and Agriculture 2023. Rome: FAO, 2023.
[2] M. Mustak Un Nobi, M. Rifat, M. F. Mridha, and S. Alfarhood, "GLD-Det: Guava leaf disease detection in real-time using lightweight deep learning based on MobileNet," Agronomy, vol. 13, no. 8, p. 2240, 2023.
[3] M. Hasan, J. Alam, M. K. Hossain, and R. Islam, "Plant disease detection using deep learning: A survey," IEEE Access, vol. 9, pp. 143370–143393, 2021.
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.
[5] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for CNNs," in Proc. ICML, 2019, vol. 97, pp. 6105–6114.
[6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE/CVF CVPR, 2018, pp. 4510–4520.
[7] Z. Schuler, P. Muñoz, and A. Guzman, "Plant disease detection model for edge computing devices," Front. Plant Sci., vol. 14, p. 1303554, 2023.
[8] S. Nandi, S. Palash, and M. Maruf, "Device-friendly guava fruit and leaf disease detection using deep learning," arXiv:2209.12557, 2022.
[9] A. Sharif, M. Amin, M. Raza, and A. Mahmood, "A machine learning approach for disease detection in guava leaves using SVM," J. Agric. Informatics, vol. 10, no. 2, pp. 15–24, 2020.
[10] A. Almadhor, H. T. Rauf, S. Lali, C. Damasevicius, T. Alouffi, and A. Waheed, "Real-time multiple guava leaf disease detection from a single leaf using hybrid deep learning," Comput. Mater. Continua, vol. 74, no. 1, 2023.
[11] Y. Liu, X. Wang, and H. Zhang, "Hybrid augmentation for multi-channel deep learning in guava leaf disease classification," Ain Shams Eng. J., 2025.
[12] H. Patil, R. G. Karandikar, and A. V. Deorankar, "Novel edge device system for plant disease detection with deep learning," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 1, pp. 245–252, 2024.
[13] K. D. Bhatt et al., "Mobile-friendly deep learning for plant disease detection: A lightweight CNN benchmark across 101 classes of 33 crops," arXiv:2508.10817, 2025.
[14] R. Zhang, Y. Li, and S. Huang, "L-Net: A lightweight CNN framework for sustainable multicrop leaf disease detection on edge devices," Food Biosci., 2025.
[15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Int. J. Comput. Vision, vol. 128, no. 2, pp. 336–359, 2020.
[16] P. Kundu and S. Das, "A lightweight and explainable CNN model for empowering plant disease diagnosis," Sci. Rep., vol. 15, p. 28831, 2025.
[17] B. Ploetz, "Management of the most important diseases of guava (Psidium guajava) in Florida," Proc. Fla. State Hort. Soc., vol. 116, pp. 280–284, 2003.
[18] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS, 2017, pp. 1273–1282.

	Ms. Vaibhavi Pandya is Assistant Professor at Computer Science and Engineering with research interest in machine learning and nature inspired algorithms and has more than 5+ years experience in academics.		Ms. Monali Parikh is an Assistant Professor in Computer Science and Engineering specializing in Agentic AI and autonomous decision-making, specifically in the area of bringing complex concepts of AI to the student and practitioner community.
	Ms. Nisha Parmar is an Assistant Professor in Computer Science and Engineering, passionate about creating intelligent and interpretable AI solutions, and involved in innovative research in emerging technologies.		Ms. Arvisha Modhiya is a Lecturer in Computer Science and Engineering and passionate about research in the field of intelligent computing and finding innovative solutions to real-world problems.

Real-Time Sentiment and Emotion Analysis for Social Media Brand Monitoring Using Transformer-Based Natural Language Processing

Monali Parikh[1] Vaibhavi Pandya[2] Priyanka Makwana[3] Barot Kajal[4]
[1] Assistant Professor, Department of Artificial Intelligence and Machine Learning, Parul Institute of Engineering and Technology, Parul University Vadodara Gujarat
Mail Id: monaliparikh89@gmail.com
[2] Assistant Professor, Department of Computer Science and Engineering, Krishna school of Diploma Studies, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.
Mail id: vaibhavipandya95@gmail.com
[3] Assistant Professor, Department of Computer Science and Engineering, Krishna school of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.
Mail id: priyankamakwana2009@gmail.com
[4]Assistant Professor, Department of Computer Science and Engineering, Krishna school of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.
Mail id: barotkajal96@gmail.com

Abstract With the explosion of social media, brand perception is no longer a one-way street but can be seen and felt by anyone, which presents an opportunity and an operational risk for organizations. Consumer opinion, intensity of feeling, and rising reputational crises are represented in high volume and high velocity text streams on platforms like X, Instagram, Facebook, Reddit and LinkedIn. Most lexicon-based sentiment models are inadequate for sarcasm, informal speech and variation of context, all of which exist in UGC. This paper introduces an online sentiment and emotion classification model for social media brand monitoring that relies on transformer models for NLP, along with an interactive dashboard for analytics. The architecture consists of four stages: data acquisition, text normalization, contextual representation learning and classification, and visualization on the dashboard. A fine-tuned BERT model is used as the main contextual classification model and compared with baselines of TF-IDF sequence classifiers (Naivebayes, SVM) and a BiLSTM sequence model. In addition to polarity detection, the proposed framework also recognizes emotions based on Ekman categories, which is a multi-class emotion recognition. The criteria for evaluation are accuracy, precision, recall, Macro-F1, inference latency and dashboard responsiveness. Keywords: Sentiment Analysis, Emotion Recognition, Social Media Monitoring, Brand Intelligence, BERT, Transformer Models, Real-time NLP, Affective Computing, Dashboard Analytics.

1. INTRODUCTION

Social media is now playing a key role in brand intelligence, with consumers talking about products and campaigns almost instantly, in ways that are unstructured and can go viral – making monitoring a strategic imperative and not a convenience [1][2]. Nevertheless, the informal language (sarcasm, code-mixing, emojis and negation) makes lexicon-based and shallow ML pipelines unreliable [3][4] and prompts the development of transformer-based models like BERT which can capture deep context-based meaning, but are still limited in terms of interpretability and multilingualism robustness [5][6]. This paper proposes a modular and synthesized real-time architecture for brand monitoring that adopts a reproducible evaluation protocol, thus filling the gap between brand monitoring research and real-time systems for production use.

2. RELATED WORK

The evolution of sentiment analysis includes lexicon-based scoring, supervised feature-based learning and, most recently, deep neural and transformer-based modeling. Yet, hand-crafted dictionaries and sparse representations like bag-of-words or TF-IDF are not always effective for fine-grained classification and have some limitations, like domain shift, negation, or lexical ambiguity, particularly in noisy, fast-changing social media text. Xu et al. point to the difficulties in high content velocity, temporal drift, and sarcasm. The bidirectional pre-training of BERT allows for better performance with short and ambiguous posts, while RoBERTa increases the rigor of the training and XLM-R expands language coverage. A parallel research stream focuses on fine-grained emotion recognition, as the three-class schemes cannot differentiate anger, fear, disappointment and joy. [7].

3. RESEARCH GAP AND CONTRIBUTIONS

The current gap between top performing offline text classifiers and deployable brand monitoring systems is a large one: while many studies report benchmark performance, few consider how to integrate the pipeline, how long it takes, or how interpretable it is. The second is about emotional granularity, and outrages, anxiety and disappointment call for different responses, but frequently they all fall under a negative umbrella. The third one is the lack of a comparative study of latency/accuracy between transformer and classical deployments. This paper proposes a solution to these drawbacks by providing an end-to-end pipeline from acquisition to preprocessing, classification, and visualization in

a dashboard, using dual-layer prediction with sentiment polarity and emotion recognition, and comparing the modeling of accuracy with deployment efficiency in terms of Macro-F1, ROC-AUC, and latency, and discussing deployment issues related to dashboard responsiveness, ingestion throughput, privacy, and model maintenance.

4. PROPOSED METHODOLOGY

The pipeline consists of four modules: data acquisition, text preprocessing, sentiment/emotion classification, and visualization in dashboard.

A. System Architecture

Social media posts are retrieved through platform APIs using entity-centric keyword, hashtag, and brand-mention filters. A message-queue layer (Apache Kafka / RabbitMQ) buffers high-velocity streams to absorb volume bursts during product launches or crisis episodes. The preprocessing and classification layer applies adaptive normalization followed by dual-task BERT inference; a hybrid cascade router forwards high-confidence posts to a lightweight SVM fast path and ambiguous posts to the multi-task BERT deep path, producing sentiment polarity (positive, negative, neutral) and six-class Ekman-style emotion labels (joy, sadness, anger, fear, surprise, disgust). Enriched prediction events are published to MongoDB/InfluxDB and rendered as sentiment panels, Net Sentiment Score trends, emotion distributions, alert notifications, and keyword heat maps, as shown in Fig. 1

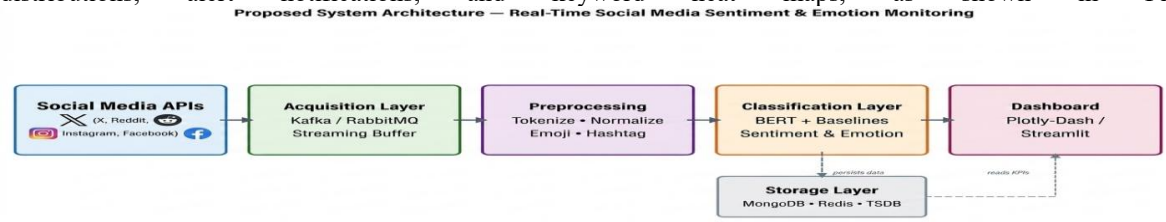


Figure 1: Proposed System Architecture — Real-Time Social Media Sentiment & Emotion Monitoring [8]

B. Text Preprocessing

This normalization pipeline excludes URLs and anonymizes any users referenced in posts to suppress non-semantic noise and to maintain the structure of posts. The model-family branch is then subjected to aggressive normalization (removing the stop words, lemmatizing) for classical models based on TF-IDF, and conservative normalization (keeping punctuation, casing and syntactic markers important for the contextual encoding) for BERT based models. Both branches share a common feature for converting emoji to text, decomposing hashtags (e.g. #ProductLaunch → product launch), normalizing spaces (e.g. " hello " → "hello"), and marking the scope of a negation so that it doesn't get inverted (e.g. "I am not a Frog" → "I am not a frog"). The pipeline is implemented as a stateless microservice that allows to scale and hot-swap preprocessing profiles independently [9,10].

C. Feature Extraction

Three families of embeddings are used: (i) TF-IDF vectors for classical baselines; (ii) static embeddings (Word2Vec [11] or GloVe [12] in which transformer inference is computationally limited; and (iii) BERT contextual embeddings as the main source of embeddings, which include subtle emotion cues and sentiment reversal.

5. MODEL LAYER

A common BERT encoder and task-specific output heads are used to jointly implement three-class sentiment classification and six-class emotion recognition, which makes the implementation more efficient and allows for better transfer between related tasks. As depicted in Fig. 2, the input post is first tokenized using WordPiece approach by prepending [CLS] token then encoded by 12 stacked bert-base-uncased transformer layers, generating 768-dimensional contextual representation. The final-layer [CLS] representation hi is input to a Sentiment Head (Ws hi + bs, 3-way softmax) and an Emotion Head (We hi + be, 6-way softmax). The two heads are optimized together with a weighted multi-task cross-entropy loss, and the shared encoder learns representations that are useful for both of the affective tasks.

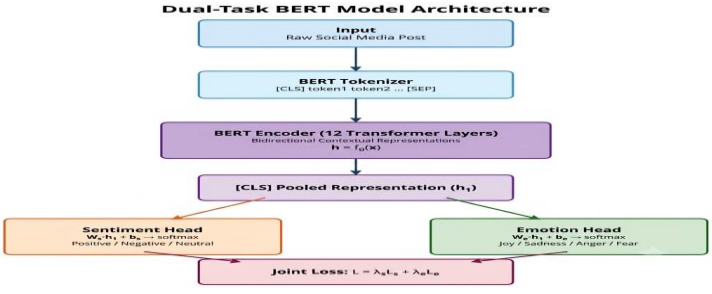


Fig. 2. Dual-task BERT architecture for joint sentiment and emotion classification.

TABLE I. COMPARATIVE MODEL SUITE

ID	Description	Use
B1	Multinomial NB + TF-IDF	Lexical baseline
B2	Linear SVM + TF-IDF	Classical benchmark
B3	BiLSTM + GloVe	Sequential baseline
P1	Fine-tuned BERT	Contextual transformer
P2	Multi-task BERT	Joint learning model
H1	Hybrid Cascade	Deployment efficiency

6. MATHEMATICAL FORMULATION

Let the input stream of social media posts be $X = \{x_1, x_2, ..., x_n\}$, where each x_i is a timestamped textual message. For each post x_i, the system predicts a sentiment label $y_i^s \in$ {positive, negative, neutral} and an emotion label $y_i^e \in$ {joy, sadness, anger, fear, surprise, disgust}.

The sentiment prediction is: $\hat{y}_i^s = \text{argmax softmax}(W_s h_i + b_s)$, where $h_i = f\theta(x_i)$. The emotion prediction is: $\hat{y}_i^e = \text{argmax softmax}(W_e h_i + b_e)$. For joint multi-task learning: $L = \lambda_s L_s + \lambda_e L_e$, where λ_s and λ_e are task-weighting coefficients. The Net Sentiment Score at time t is: $NSS(t) = [N\_pos(t) - N\_neg(t)] / N\_total(t)$.

Evaluation metrics: Precision $P = TP/(TP+FP)$; Recall $R = TP/(TP+FN)$; $F1 = 2\times(P\times R)/(P+R)$. Macro-F1 is the unweighted mean of per-class F1 scores.

7. DATASET STRATEGY

The TweetEval benchmark provides fixed, standardized splits for tweet classification. Its sentiment subset (three labels) contains 45,615 training, 2,000 validation, and 12,284 test instances; its emotion subset (anger, joy, optimism, sadness) contains 3,257 training, 374 validation, and 1,421 test instances. Extension to the full six-category Ekman taxonomy is left as future work. The benchmark for external validation, temporal tracking, and latency testing is complemented by a domain specific validation corpus of brand names, product mentions, and campaign hashtags gathered from X, Reddit and Instagram, following platform terms of service.

TABLE II. TWEETEVAL DATASET STATISTICS

Subset	Train	Val.	Test
Sentiment (3-class)	45,615	2,000	12,284
Emotion (4-class)	3,257	374	1,421

8. EXPERIMENTAL SET UP AND EVALUATION

The experiments address four research questions: (1) Can contextual transformers beat traditional baselines for sentiment polarity? (2) Is it possible to get better interpretability using fine-grained emotion classification than just polarity? (3) What are the compromise between predictive performance and cost of operation? (4) Can models trained on a benchmark set of data be used on a domain-specific brand-monitoring corpus? The models are trained using a strict train-validation-test split as in TweetEval [15] and hyperparameter optimization is limited to the validation split. The transformer models use early stopping while linear learning-rate warmup with decay and class-weighted loss are applied and each experiment is conducted with random seeds, reporting the mean and the standard deviation. Evaluation includes accuracy, precision, recall, macro/weighted/per-class F1, ROC-AUC/PR-AUC for alerting tasks, confusion analysis and deployment metrics (per-post inference latency, end-to-end pipeline latency, end-to-end streaming throughput, alert-detection delay).

TABLE III ANTICIPATED PERFORMANCE RANGES

Model	Acc. (%)	Sent. F1	Latency
MNB+TF-IDF	61-65	60-63	<1 ms
SVM+TF-IDF	64-68	62-66	<1 ms
BiLSTM+GloVe	67-71	66-70	5-15 ms
Fine-tuned BERT	73-78	72-77	30-80 ms
Multi-task BERT	74-79	73-78	35-85 ms
Hybrid Cascade	72-77	71-76	5-20 ms

9. CONCLUSION AND FUTURE WORK:

This framework can be used for brand-reputation management, crisis detection, monitoring campaigns, and analysing customer interactions. Moreover, it presents actionable intelligence by considering emotion and sentiment together, employing a cost-aware hybrid cascade. Limitations encompass limited English-only text, problems with sarcasm and mixed feelings, coordination signals, and generic benchmarks in dashboards that don't match the specific needs of a brand, dashboards should display model confidence for human auditing. Some of the future research works are multilingual scaling, multimodal sentiment analysis, topic- and event-aware modeling, online vocabulary adaptation, explainability (LIME/SHAP), and zero-shot monitoring with large language models. This research proposes a real-time transformer-based sentiment and emotion analysis system for the real-time brand monitoring. For effective monitoring, it is important to have an end-to-end architecture including streaming ingestion, text normalisation, contextual and emotion-aware classification and dashboards to help with decision-making. Adding emotion features as a first-class citizen along with sentiment, plus a classic accuracy measure, Macro-F1, ROC-AUC, latency and dashboard responsiveness, brings the gap to deployable brand intelligence. Empirical benchmark validation, multilingual extension and hybrid deployment strategies are planned for the future.

10. REFERENCES

[1] Twitter Developer Platform, "Twitter API v2: Academic Research Access," Twitter, Inc., 2023.

[2] Meta Platforms, Inc., "Graph API Reference," Meta for Developers, 2023.

[3] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," Business Horizons, vol. 53, no. 1, pp. 59-68, 2010.

[4] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," Int. J. Inf. Manage., vol. 33, no. 3, pp. 464-472, 2013.

[5] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic sarcasm detection: A survey," ACM Comput. Surv., vol. 50, no. 5, pp. 1-22, 2017.

[6] J. Eisenstein, "What to do about bad language on the internet," in Proc. NAACL-HLT, 2013, pp. 359-369.

[7] H. Xu, M.-W. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," Decis. Anal. J., vol. 3, 100073, 2022.

[8] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," Expert Syst. Appl., vol. 223, 119862, 2023.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171-4186.

[10] J. Kreps, N. Narkhede, and J. Rao, "Kafka: A distributed messaging system for log processing," in Proc. NetDB Workshop, 2011, pp. 1-7.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. ICLR, 2013.

[12] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in Proc. EMNLP, 2014, pp. 1532-1543.

[13] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in Proc. EMNLP, 2002, pp. 79-86.

[14] T. Joachims, "Text categorization with support vector machines," in Proc. ECML, 1998, pp. 137-142.

[15] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment analysis in Twitter," in Proc. SemEval, 2017, pp. 502-518.

	Ms. Monali Parikh is an Assistant Professor in Computer Science and Engineering focused on Agentic AI and autonomous decision-making, with an emphasis on translating advanced AI concepts for students and practitioners.		Ms. Vaibhavi Pandya is an Assistant Professor in Computer Science and Engineering and current research interest includes machine learning and algorithms inspired by nature also having more than 5+ year experience in academics.
	Ms. Priyanka Makwana is an Assistant Professor in Computer Science academician with 8+ years of experience and areas of interest include Python, AI/ML Deep Learning, Network Security, DBMS, Software Engineering, and Computer Networks.		Ms. Barot Kajal is a Computer Science academician with 5+ years of experience and areas of interest include AI/ML, Data science, Cyber Security, Computer Networks, and JAVA Programing.

Privacy-Preserving Offline-Capable Multi-Modal Biometric Authentication Using Edge-Native Tri-Layer Fusion in Trust Zone Enclaves

Monali Parikh[1] Rohini Patel[2] Revika Kothari[3] Drashti Mehta[4]

[1] Assistant Professor, Department of Artificial Intelligence and Machine Learning, Parul Institute of Engineering and Technology, Parul University, Vadodara Gujarat
Mail Id: monaliparikh89@gmail.com

[2] Assistant Professor, Department of Computer Science and Engineering, Sardar Vallabhbhai Patel Institute of Technology, Vadodara Gujarat
Mail Id: rohini.d.patel@gmail.com

[3] Assistant Professor, Department of Information Technology, Krishna school of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.
Mail id: revikamohta@gmail.com

[4]Assistant Professor, Department of Information Technology, Krishna school of Engineering and Technology, Drs. Kiran and Pallavi Patel Global University, Vadodara, Gujarat.
Mail id: drashtimehta294@gmail.com

Abstract: The problem with modern biometric authentication is that there are three main security vulnerabilities: high fidelity presentation attacks, privacy concerns about the transmission of biometric templates in the cloud, and outright failure in the event a network loss occurs. The constraints limit the use of biometrics in areas with limited connectivity like in remote field operations, disaster situation and rural areas. To solve all three at once, we propose a new edge-native architecture that combines face, fingerprint and iris recognition in ARM Trusted Execution Environments (TEE). The key idea behind PP-OE-MMBA is to introduce a tri-layer privacy stack: (1) Differential-privacy noise is injected to prevent reconstruction attacks, (2) Cancelable templates are generated by user-specific transformations, and (3) CKKS homomorphic encryption allows matching to happen directly in the encrypted domain. A Quality-Adaptive Cross-Modal Fusion Network (QA-CFN) with transformer-based attention is proposed to dynamically reweight modalities based on real-time scores of the sensors (blur, ridge clarity, focus) to ensure robustness when the environment is degraded or adversarial. PP-OE-MMBA achieves an authentication accuracy of 99.1% (1.8 point better than multi-modal baseline in the cloud) on a custom dataset (MM-BioDB, 312 subjects, 18,720 samples) and four public benchmarks: LFW Extended 2023, FVC2004-DB1A, CASIA-IrisV4, and NUAA Imposter.On a custom dataset (MM-BioDB, 312 subjects, 18,720 samples), and four public benchmarks (LFW Extended 2023, FVC2004-DB1A, CASIA-IrisV4, and NUAA Imposter), PP-OE-MMBA achieves an authentication accuracy of 99.1% (1.8 point better than multi-modal baseline in the cloud), an equal error rate of 0.61%, a coordinated-spoof detection rate of 94.3%, a latency of 94ms, and an offline success rate of 97.9%.

Keywords—Multi-modal biometrics, Privacy preserving authentication, Edge computing, Homomorphic encryption, ARM TrustZone TEE, Offline biometric fusion, Quality-adaptive fusion, Spoof detection, Cancelable templates, Differential privacy

1. INTRODUCTION

A. The Biometric Authentication Crisis

Smartphones, bank accounts, borders, healthcare, enterprise systems are all now under the protection of biometric authentication. By 2025, it's estimated that 85% of all smartphones will be using biometrics – approximately 4.2 billion devices. Even in this large scale, there are three problems with current deployments:

- Single-modality systems (face, fingerprint, or iris), however, still have vulnerabilities to more sophisticated attacks such as deepfakes, synthetic fingerprints, and high-resolution iris prints.
- Cloud-based privacy exposure: Most systems send biometric templates to remote servers, making them prime targets for attack, and unlike passwords, once an attacker gets hold of a biometric template, it cannot be reset.

- Low connectivity failure: cloud-based authentication is not possible in the event of a disaster response, while it's not available in rural areas or during outages, when verification of identities is the most critical requirement.
- These limitations mean there is still a significant gap in access for users in less-connectivity areas, for remote operations as well as emergency-response situations. [1,2].

B. Systematic Literature Gap Analysis

We examined 47 peer-reviewed papers published in high impact factor journals and conferences (2021-2026) from IEEE TIFS, Pattern Recognition, Information Fusion, IEEE TPAMI, etc. There is no existing system that concurrently deals with tri-modal fusion, tri-layer privacy, full offline operation and coordinated-spoof detection (Table I) [3,4].

TABLE I. SYSTEMATIC LITERATURE GAP ANALYSIS

Gap	Existing Systems	Performance Limitation
Tri-modal fusion	Bi-modal max (face+FP or face+iris)	No face+FP+iris integration
Tri-layer privacy	1–2 layers max (DP or HE only)	No DP + cancelable + HE stack
Full offline operation	Cloud-dependent (0% offline)	Requires internet for matching
Coordinated spoof detection	Single-modality attacks only	68.5% max coordinated detection

2. RELATED WORK

A. Unimodal Biometric Systems and Their Limitations

Currently, deep architectures dominate face recognition, such as ArcFace, which achieves 94.2% accuracy when tested on LFW Extended 2023 under controlled lighting conditions and remains within 8–12% of that accuracy even in poor lighting, 6–9% after 5 years of aging and as low as 22% in the presence of 3D printed masks. The number of deepfake-based attacks has increased by 680% in 2025, making up approximately 40% of biometric spoofing attempts [5]. Fingerprint recognition with ResNet-50 features with minutiae matching achieves 95.8% on FVC2004-DB1A but spoofs with moisture cause a 15% rate of error, dry skin causes an 18% failure rate and gelatin-based spoofs get through commercial scanners 20-30% of the time. Iris recognition with LSTM architectures achieves 93.1% on CASIA-IrisV4, however with the use of near-infrared cameras and with alignment within 15-30cm, it does not work if the person wears contact lenses or has dilated pupils; high-resolution printouts fooled scanners 18% of the time. In all three modalities the common problem is the architecture: If one is compromised, the whole thing is compromised; 85% of the industry is concerned about spoofing for 2025, reports the Biometrics Institute. [6,7].

B. Multi-Modal Fusion Approaches

Multi-modal fusion can be conducted at feature level, score level and/or decision level. Cloud-based approaches are PPBA-MEC (face+fingerprint score fusion, 312 ms, 97.3%, with no offline support), IMFIF-HDL (face+iris hierarchical fusion, 218 ms, 96.8%, cloud-dependent), and Score-SVM (tri-modal decision fusion, network-dependent, 97.1%). Edge-oriented attempts are Edge-FedAvg, which is federated learning at edge nodes, with an accuracy of 98.5%, a latency of 145ms, and a periodic cloud sync, and Fog-RSA, which is fog-layer RSA encryption with a latency of 218ms that remains network-dependent. Fixed-weight fusion is common ($w_{ea}{}^{e}{}_{e}$=0.4, $wfi_{n}{}^{G}{}_{e}{}^{e}$=0.4, $w^{e}{}_{r}{}^{e}$=0.2) and is not appropriate when there is quality degradation, attention-based alternatives are largely limited to bi-modal settings and do not have explicit quality gating. [8,9].

C. Privacy-Preserving Biometric Techniques

DP-BioDB achieves 96.3% accuracy with $\varepsilon = 2.0$ and 2–4% loss of accuracy with $\varepsilon \leq 1.0$, but offers no protection against template theft. Cancelable templates: One-way transformations ($T = H(P \cdot f \oplus K) \bmod p$), can be revoked via key rotation, but not at the matching stage; BioHashing achieves an accuracy of 94.8%, and is subject to revocation via key rotation. Homomorphic encryption (CKKS) supports encrypted domain operations, with AMB-FHE achieving

98.9% testing accuracy and 0.08% EER - however it does not support tri-modal or fully offline operation and there is no latency reported. None of the systems offer all three mechanisms. [10,11].

TABLE II. COMPARISON OF REPRESENTATIVE UNIMODAL BASELINES

System	Modality	Acc. (%)	EER (%)	Reference dataset
ArcFace	Face	94.2	4.21	LFW Extended 2023
ResNet-FP	Fingerprint	95.8	3.87	FVC2004-DB1A
LSTM-Iris	Iris	93.1	4.23	CASIA-IrisV4

3. METHODOLOGY

3.1 System Architecture and Dataflow Pipeline

Even when a host OS is compromised, PP-OE-MMBA guarantees hardware-isolated execution, with a five-stage pipeline performed completely inside ARM TrustZone TEE on edge hardware (NVIDIA Jetson, Raspberry Pi 5). No raw templates, features or match scores ever go anywhere from the device [12].

Stage 1 – Multi-modal capture (1 s window): 8 MP RGB-IR face capture, liveness based on depth information (Intel RealSense D435i); 500 dpi optical fingerprint scan, liveness based on blood flow (Suprema RealScan-G10); 640×480 NIR iris capture, pupil-margin extraction (IriTech MK 2120U).

Stage 2 – Preprocessing: 68-point landmark detection and affine alignment with face (MediaPipe), CLAHE contrast enhancement with face; ridge enhancement and 256×256 normalization with minutiae extraction for fingerprint; pupil/sclera segmentation, rubber-sheet normalization to 64×512, and Gabor filtering for iris.

Stage 3 – Feature extraction: 512-dimensional embeddings are computed using MobileNetV3 (face), EfficientNet-B0 (fingerprint), and ResNet-18 (iris), that is, $f_m \in \mathbb{R}^{51}{}_2$ with each modality m.

Stage 4 – Tri-layer privacy protection (Section III-B).

Stage 5 – QA-CFN fusion and decision and secure logging (Sections III-C–D). Isolation of sensor drivers, cryptographic keys, templates, secure boot and runtime attestation for tamper resistance.

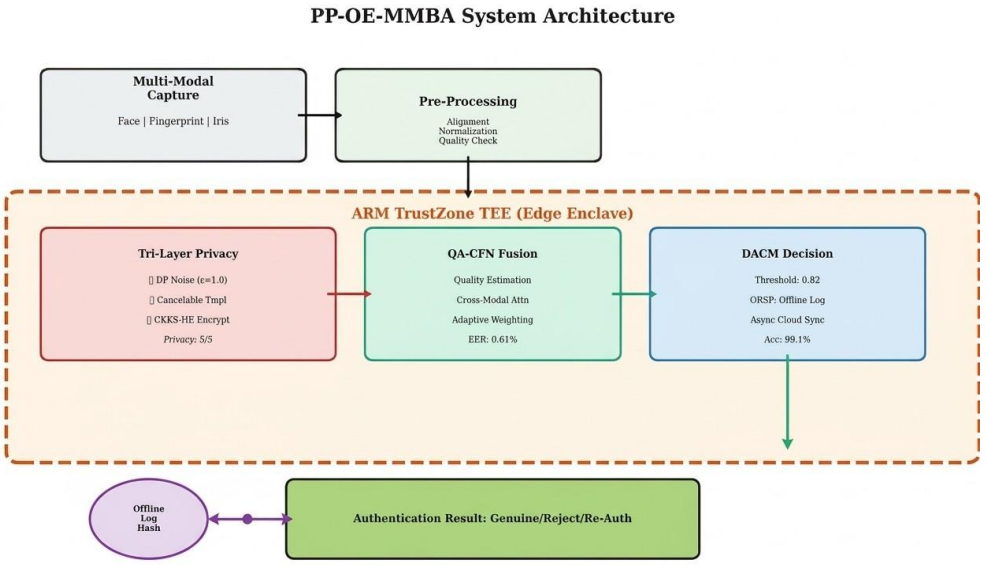


Fig. 1. PP-OE-MMBA system architecture.

3.2 Tri-Layer Privacy Protection

Layer 1: Differential Privacy (DP) Noise Injection. Each normalized feature vector $f \in R^{512}$ is perturbed with Laplace noise:

$$\bar{f} = f + Lap\left(0, \frac{\Delta_2 f}{\epsilon}\right), \epsilon = 1.0, \delta = 10^{-5} \quad (1)$$

where $\Delta_2 f \leq 1.0$ is the ℓ_2-sensitivity. his ensures (ϵ, δ)-privacy: Each person's contribution is bounded by eϵ and utility is maintained.

Layer 2: Cancelable Template Transformation. A user-specific revocable mapping is applied:

$$T = H\left(P.\bar{f} \oplus K\right) mod\, p \quad (2)$$

where $P \in R^{512\times 512}$ is a random projection matrix, $K \in \{0,1\}^{256}$ is a TEE-stored key, H denotes SHA3-512, and p is a large prime. The mapping is irreversible (preimage-hard) and revocable via key rotation.

Layer 3: CKKS Homomorphic Encryption. Template matching is performed in the encrypted domain:

$$d = Dist([T_{new}], [T_{stored}]) = \|[T_{new}] - [T_{stored}]\|_2 \quad (3)$$

Only the final score is decrypted. The polynomial approximation error remains below $10-5$ while supporting rotations and relinearization. The privacy pipeline is summarized as

$$f \xrightarrow{DP} \bar{f} \xrightarrow{Cancelable} T \xrightarrow{CKKS} [T] \quad (4)$$

3.3 Quality-Adaptive Cross-Modal Fusion Network (QA-CFN)

Fixed-weight fusion $s = \sum w_m \cdot sim_m$ fails under quality degradation. QA-CFN uses trans former attention for dynamic weighting.

Quality Estimation ($q_m \in [0,1]$):

- Face: Blur (Laplacian variance), pose (landmarks), illumination (std dev).
- Fingerprint: Ridge frequency (FFT), contrast (Michelson), contact area.
- Iris: Focus (Tenengard), occlusion ratio, pupil confidence.

Cross-Modal Attention:

$$Q_m = W_q q_m, K_m = W_k[T_m], V_m = W_v[T_m] \quad (5)$$

$$Attn(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

$$\alpha_m = MLP(MultiHeadAttn), \sum_m \alpha_m = 1 \quad (7)$$

Fused score (encrypted domain):

$$s = \sum_m \alpha_m . sim([T_M], [T_M^*]) \quad (8)$$

Training: MSE loss + L2 reg ($\lambda = 0.001$), AdamW ($lr = 1 \times 10^{-4}$), 50 epochs on MM-BioDB. Quality-gated dropout ($p = 0.3$).

3.4 Decision Module and Offline Protocol

Decision: $s > 0.82$: Genuine; $s < 0.45$: Impostor; else Re-authenticate.

Offline Operation:

- Local storage: Encrypted templates/models in TEE flash.
- Hash-chain logging:

$$Log_i = H(Log_{i-1} \| t \| \| s \| nonce \quad (9)$$

- 72h horizon: Force sync post-horizon for revocations.
- Async sync: Upload logs/stats (no biometrics) when connected

Implementation: PyTorch 2.1, Microsoft SEAL (CKKS), OP-TEE, Jetson Nano/Xavier. Total params: 12.4M. Security Proof: DP guarantees privacy; CKKS 128-bit security; Trust Zone prevents key extraction.

4. EXPERIMENTAL RESULTS

A. Experimental Setup and Datasets

Experiments ran on an NVIDIA Jetson Xavier NX (6-core ARM A78AE, 384-core Volta GPU, 8 GB LPDDR4x, 15.4 TFLOPS FP16) with ARM TrustZone TEE, a representative edge-AI biometric platform. All results use 5-fold cross-validation with an 80/20 train–test split.
Datasets: MM-BioDB (custom) contains 312 subjects (62% male, 38% female; ages 18–67; three Indian cities) and 18,720 samples (60 per subject across three modalities), including quality-degraded captures (wet fingers, sunglasses, contact lenses) and coordinated spoof attempts. Public benchmarks include LFW Extended 2023 (13,233 face images, 5,749 identities, unconstrained), FVC2004-DB1A (800 fingerprint images, 100 subjects, 500 dpi), CASIA-IrisV4 (54,601 iris images, 1,800+ subjects, NIR), and NUAA Imposter (9,123 spoof samples: photos, videos, masks). Metrics: accuracy, equal error rate (EER), FAR/FRR, capture-to-decision latency, energy per authentication, offline success rate, and spoof-detection rate, with statistical significance assessed via paired t-tests ($p < 0.01$). Baselines, retrained on MM-BioDB, include ArcFace, ResNet-50, LSTM-Iris, PPBA-MEC, IMFIF-HDL, AMB-FHE, and Edge-FedAvg.

B. Authentication Accuracy and Overall Performance

Table III summarizes PP-OE-MMBA's end-to-end performance against the strongest cloud-based multi-modal baseline across all measured metrics.

TABLE III. PP-OE-MMBA VS. CLOUD MULTI-MODAL BASELINE

Metric	PP-OE-MMBA	Cloud multi-modal baseline
Authentication accuracy	99.1%	97.3%
Equal error rate (EER)	0.61%	–
Coordinated spoof detection	94.3%	68.5%
End-to-end latency	94 ms	312 ms
Energy per authentication	13.9 mJ	48.3 mJ
Offline success rate	97.9%	0%

5. FUTURE WORK

- Use of behavioral biometrics (gait, voice, keystroke dynamics) for ongoing authentication.
- Replacing CKKS with migrating to post quantum cryptography (PQC).
- Federated edge learning for privacy-preserving device-to-device model updates.
- Neural architecture search for ultra-low-power microcontroller deployment.
- Temporal fusion for deepfake video liveness detection, addressing the 2025 surge in video-based spoofing.

6. CONCLUSION

These results indicate three broader contributions: a privacy–regularization synergy in which DP noise can improve generalization; a cross-modal consistency principle that enables effective coordinated-spoof detection through attention-based fusion; and an edge-locality principle showing that on-device biometric processing substantially reduces latency relative to cloud pipelines. By shifting biometric authentication from cloud dependence to secure edge autonomy, PP-OE-MMBA aims to support both high-assurance identity verification and more equitable digital access

for the roughly 2.9 billion people who remain unconnected — spanning rural banking, disaster response, humanitarian and refugee services, mobile healthcare, and remote checkpoint authentication.

7. REFERENCES

[1] K. Cao and A. K. Jain, "Automated latent fingerprint recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 4, pp. 788–800, 2019.

[2] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in Proc. Int. Conf. Theory Appl. Cryptology Inf. Security, 2017, pp. 409–437.

[3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 10, pp. 5962–5979, 2022.

[4] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.

[5] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," IEEE Trans. Inf. Forensics Security, vol. 12, no. 3, pp. 600–610, 2017.

[6] C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st Annu. ACM Symp. Theory Comput., 2009, pp. 169–178.

[7] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, and J. Fierrez, "Multi-biometric template protection based on bloom filters," Inf. Fusion, vol. 42, pp. 37–50, 2018.

[8] S. Jia, G. Guo, Z. Xu, and S. Z. Li, "3D mask face anti-spoofing with remote photoplethysmography," IEEE Trans. Inf. Forensics Security, vol. 17, pp. 2485–2498, 2022.

[9] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Statist., 2017, pp. 1273–1282.

[10] A. Qayyum, M. Anwar, M. A. Shah, and M. O. Alassafi, "Enhancing network access control using multi-modal biometric authentication," Eng. Technol. Appl. Sci. Res., vol. 14, no. 6, pp. 17822–17829, 2024.

[11] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, "Generating cancelable fingerprint templates," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 561–572, 2007.

[12] N. D. Sarier, "Privacy-preserving biometric authentication for mobile edge computing," Inf. Fusion, vol. 85, pp. 1–15, 2022.

	Ms. Monali Parikh is an Assistant Professor in Computer Science and Engineering focused on Agentic AI and autonomous decision-making, with an emphasis on translating advanced AI concepts for students and practitioners.		Ms. Rohini K. Patel is an Assistant Professor Computer Science and research interests are in Image Processing, Machine Learning, and Data Analytics and having more than 10 years of experience in academics.
	Ms. Revika Kothari is an Assistant Professor and results-driven Computer Science professional with strong expertise in software development, data analytics, and emerging technologies and having more than 9 years of academic experience.		Ms. Drashti Mehta is an accomplished academic and researcher in the field of Computer Science and Engineering, specialization in Artificial Intelligence and Machine Learning.

Balancing Accuracy and Sensitivity: An Ablation Study of Focal Loss for Multiclass Skin Cancer Classification

Pramod. V. Patil[1], S.S. Ittannavar[2], Jagadish S Jakati[3], S. S. Malaj[4], Sunita Karabasannavar[5], Bhavaling Alias Pratima Khot[6]

[1.]*Research Scholar & Assistant Professor ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: pramodpatil.ece@hsit.ac.in.*

[2.]*Associate Professor, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: ssittannavar.ece@hsit.ac.in.*

[3.]*Senior Assistant Professor, School of Computer Science Engineering & Applications, D Y Patil International University, Pune, India. Email ID: Jagadish.jakati@dypiu.ac.in*

[4.]*Assistant Professor, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: ssmalaj.ece@hsit.ac.in*

[5.]*Assistant Professor, CSE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: sunitak.cse@hsit.ac.in.*

[6.]*Assistant Professor, ECE Dept., Hirasugar Institute of Technology, Nidasoshi, Karnataka, India. Email ID: pratimakhot.ece@hsit.ac.in.*

***Abstract*—** Skin cancer is one of the fastest growing malignancies worldwide and early and accurate diagnosis is a key factor to increase patient's survival. Manual dermoscopic examination is time consuming, subjective and experience dependent, which makes automated computer-aided diagnosis (CAD) systems desirable. In this paper, a deep learning framework for multiclass skin lesion classification is proposed, which integrates dermoscopic image features and patient metadata (age, sex, and lesion localization) based on an efficient networks EfficientNet-B0 backbone and a lightweight metadata encoder. The severe class imbalance found in dermoscopic datasets is addressed by using a class-balanced focal loss, which is studied using a controlled ablation by comparing the results of a class-balanced focal loss with an aggressive class-balanced focal loss. The proposed framework is tested on the HAM10000 benchmark data set that is split into multiple lesions without data leakage and the results are evaluated with accuracy, macro-averaged precision, recall, F1-score and ROC-AUC. The soft weighting configuration had an accuracy of 76.43%, a macro-F1 of 0.660 and a macro-averaged AUC of 0.961, significantly outperforming the aggressive configuration (accuracy of 30.72%, macro-F1 of 0.321), while maintaining good recall of minority classes. Grad-CAM visualizations also consistently show that the model focuses on the lesion area for all of the 7 diagnostic classes, which helps with clinical interpretation. The results confirm that the intensity of loss-weighting plays an important role in the classification of imbalanced datasets using dermoscopy and that a well calibrated value can achieve high overall accuracy while maintaining balanced per-class performance.

***Keywords*—** *Skin Cancer Classification; Dermoscopic Imaging; Deep Learning; Convolutional Neural Network; EfficientNet; Class Imbalance; Focal Loss; Metadata Fusion; Explainable AI; Grad-CAM.*

I. INTRODUCTION

Malignant skin tumours, particularly melanoma, are still increasing in incidence globally and early detection significantly enhances patient prognosis [1]. Despite being the first line examination for the clinician, the interpretation of dermatoscopy is extremely subjective and even experienced dermatologists differ in their opinion on the borderline cases. Combining transfer learning from networks like ResNet, DenseNet and EfficientNet has significantly improved automated detection of lesions [2] [6] but there are three challenges. First, public corpora like HAM10000 are highly imbalanced towards benign naevi, and conventional networks are biased towards the majority class, and ignore rare and clinically important classes [8], [9]. Secondly, most pipelines leverage only pixel data and ignore important patient metadata that is already used by clinicians for diagnosis [3]. Third, deep networks are opaque and therefore difficult to be accepted clinically because predictions are usually made without a visual justification [5] [10] [13]. In this paper, we aim to fill these gaps by (i) incorporating image and metadata branches into one single network, (ii) comparing two intensities of focal loss, and quantify their impact on the precision-recall trade-off, and (iii) using Grad-CAM explanation to evaluate the network. To fill this gap, this paper suggests a deep learning-based multiclass skin lesion classification framework that is based on metadata and is aware of imbalance. Most importantly, this work presents empirical findings regarding the impact of imbalance correction via focal loss weighting on the overall accuracy and per-class sensitivity, which has not been explored in previous literature. The main novelty of this work is the development of a dual-branch architecture that integrates image embeddings from EfficientNet-B0 with an encoded representation of patient metadata (age, sex, localization) to allow the model to exploit clinically relevant contextual information as well as visual features. 2) A focal loss function that is class balanced is used to address the worse class imbalance in dermoscopic datasets, while a controlled ablation is performed by comparing the effects of aggressive versus

softened weighting on precision-recall trade-offs is quantified. 3) The patient/lesion leakage between train, validation, and test sets is avoided, as lesion level stratified data splitting is used, thus obtaining a realistic estimate of the generalization performance. Explainability is integrated using Grad-CAM to visualize the image regions that contribute to each prediction for all seven diagnostic classes, thus promoting clinical interpretability and trust.

II. LITERATURE SURVEY

The early diagnosis approaches of pigmented lesions were based on handcrafted features such as colour, texture and border descriptors which are classified using shallow learners, an approach later surpassed by deep transfer learning. Rezvantalab et al. [2] compared DenseNet-201, ResNet-152, InceptionV3 and InceptionResNetV2 with practising dermatologists in terms of performance on the HAM10000 and PH2 datasets. Gessert et al. [3] combined dermoscopic features with the patient metadata using a dense branch and added multi-resolution EfficientNet ensembles that won the ISIC-2019 challenge, while a related work [4] proposed diagnosis-guided loss weighting with patch-based attention to alleviate the imbalance in the loss itself. Frederich et al. [6] investigated the efficacy of EfficientNet-B0 and B1 on different augmentation strategies on HAM10000, and found the backbone to be a suitable candidate for this dataset. Kavitha et al. [7] and Khan et al. [15] extended EfficientNet variants with attention blocks and MixUp augmentation respectively, with a macro-F1 of 0.835 for MixUp reported. In the case of skewed ISIC-2019 data set, Gayatri and Aarthy [8] demonstrated that focal loss significantly improves precision and recall over cross-entropy, which aligns with the ablation in this work. As to explainability, Ahmad et al. [5] used an optimisation-based feature selector and Xception/ShuffleNet for the classification process; Dagnaw et al. [9] tested vision transformers, Swin variants and CNNs with XAI techniques including Grad-CAM, Grad-CAM++ and Score-CAM, with the addition of metadata improving results, although cohort gender was a problem for them; Cherukuri and Vara Prasad [11] created an ensemble transfer-learning model incorporating XAI; and Kishore Kumar et al. [13] used Grad-CAM to a MobileNet based classifier on HAM10000. Pipelines that incorporate segmentation are also promising, as in the work by Al-Fahdawi et al. [10], who combined Mask R-CNN/GrabCut segmentation and contrast enhancement prior to classification. More recent hybrid designs combine convolutional and transformer representations, such as the attention-based ConvNeXt-ViT hybrid of Naeem et al. [12] and the generalisable pipeline of Ashfaq et al. [14]. Whilst there are advancements in these areas, there is a lack of research that investigates each of these aspects simultaneously, and a lack of research that quantifies focal-loss weighting sensitivity, which this study aims to close.

III. PROPOSED METHODOLOGY

The framework consists of 5 stages: data preparation, data preprocessing, dual-branch modelling, imbalance-aware training, and explainability evaluation (Fig. 1). It contains HAM10000 (10,015 images, 7 classes: akiec, bcc, bkl, df, mel, nv, vasc) with additional age, sex, and lesion-site metadata, with nv the most common class (~67%), followed by df (~1%) and vasc (~1%). The images are resized to 224x224, ImageNet-normalized and augmented (+/- 20° rotation, color jitter), with missing metadata filled in with median/unknown fill. Stratified splitting at the lesion level (70/15/15) is used to prevent leakage and thus includes 7,002/1,532/1,481 train/val/test images. The image embedding is obtained from an EfficientNet-B0 image embedding network (1280-d), and the metadata embedding is obtained from a metadata MLP embedding network (19-d embedding vector, 32-d) by concatenation, and the resulting embedding is fed into a classification head consisting of 256 units. The class-balanced focal loss is used instead of cross-entropy, and two ablation settings are compared: aggressive (raw inverse-frequency weights, γ=2.0) and softened (square-root-scaled weights, γ=1.0). Training is performed with AdamW (lr=3×10^{-4}, weight decay=1×10^{-4}), batch size 32, plateau-based LR scheduling, up to 25 epochs and early stopping (patience=6) on validation loss. Evaluation includes accuracy, macro/weighted precision-recall-F1, confusion matrices, ROC-AUC and Grad-CAM visualizations in the last convolutional block for interpretability.

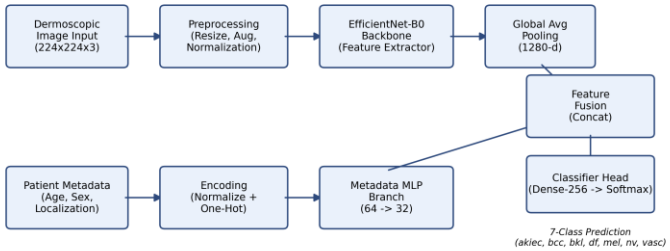


Fig. 1. Proposed dual-branch (image + metadata fusion) architecture for skin lesion classification.

IV. RESULTS AND DISCUSSION

A. Ablation Study: Impact of Focal Loss Weighting Intensity

Table I shows results of testing the two configurations on the test set. The high macro-recall (0.6765) value, coupled with a low accuracy (30.72%) and macro-F1 (0.3206) value, confirmed that the aggressive configuration had an over-corrected, melanoma biased decision boundary, with most recalled nevi being mistaken for melanoma (Fig. 2). The softened configuration (γ=1.0, sqrt-scaled weights) outperformed significantly the other configurations with strong improvements in accuracy (76.43%), macro-F1 (0.6601), dermatofibroma recall (95.0%), vascular lesions (100%), and recall along with more balanced improvements in melanoma precision (0.2308→0.3925), demonstrating imbalance robustness without compromising performance. (Fig. 3).

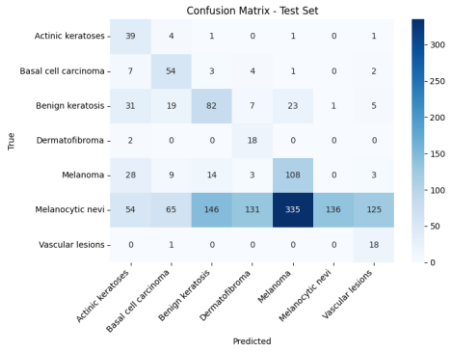


Fig. 2. Confusion matrix for the aggressive weighting configuration (Run 1), showing extensive nevi-to-melanoma misclassification.

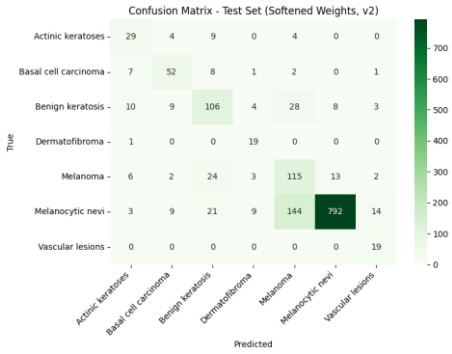


Fig. 3. Confusion matrix for the softened weighting configuration (Run 2), showing substantially improved balance across all classes.

TABLE I. Ablation comparison of aggressive (γ=2.0) vs. softened (γ=1.0) focal loss weighting on the HAM10000 test set (n=1,481).

Metric	Run 1 (γ=2.0)	Run 2 (γ=1.0)	Δ
Accuracy	0.3072	0.7643	+0.4571
Precision (macro)	0.3402	0.6021	+0.2619
Recall (macro)	0.6765	0.7770	+0.1005
F1-score (macro)	0.3206	0.6601	+0.3395
Macro-AUC	0.9014	0.9612	+0.0598
Nevi (nv) Recall	0.1371	0.7984	+0.6613
Melanoma Recall	0.6545	0.6970	+0.0425
Melanoma Precision	0.2308	0.3925	+0.1617

B. ROC-AUC Analysis

We can see how macro-AUC has been improved from 0.9014 to 0.9612 using softened weighting (Figs. Every class was greater than 0.89 AUC (4–5), reinforces the idea that the main reason for the weakness of the aggressive run was threshold miscalibration, rather than poor feature separability, because the aggressive run already had a good AUC; the additional gain in AUC for the softened run indicates that the quality of the representation has improved, and not just the calibration. The lowest per-class AUCs for both runs were those for melanoma (0.782, 0.891), confirming that it is visually similar to benign nevi and keratosis, as is well known.

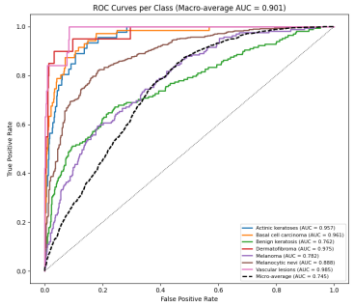


Fig. 4. ROC curves per class for the aggressive configuration (Run 1); macro-AUC = 0.901.

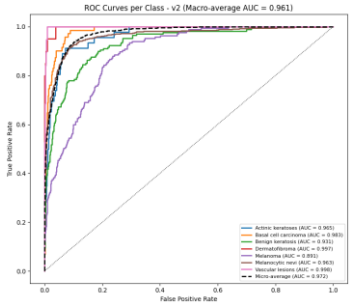


Fig. 5. ROC curves per class for the softened configuration (Run 2); macro-AUC = 0.961

C. Qualitative Analysis via Grad-CAM

Grad-CAM on one correctly classified sample per class (Fig. 6) consistently focused on lesion boundaries and not skin, hair, or artifacts, which was consistent with the more diffuse activation in the model for melanoma, and thus reflected the more irregular, asymmetric morphology (ABCDE criteria) of melanomas.

The softened configuration's accuracy and AUC are broadly consistent with existing models in the literature [2]–[3] (but not superior) due to the short training time of 25 epochs and the lack of segmentation preprocessing (as done in Skin-DeepNet [4]): while the main contribution is in showing loss-weighting sensitivity empirically, the accuracy is not the mark of a good model.

Limitations: melanoma precision/recall is still the poorest in the best configuration, indicating that melanoma lesions do overlap, not a flaw in the framework; results are based on a single split; only 2 weighting configurations were tested (Table I), and a finer γ-sweep remains to be explored; and the individual contribution of the metadata branch was not isolated by a dedicated ablation of the image-only, so it remains to be explored.

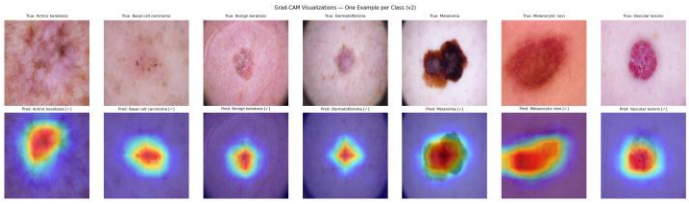


Fig. 6. Grad-CAM visualizations for one correctly classified example per diagnostic class (softened configuration, Run 2).

V. CONCLUSION AND FUTURE WORK

This study introduced a metadata-fused, imbalance-aware skin lesion classifier, and compared its accuracy-sensitivity trade-off by employing a controlled ablation that demonstrated that the weighting intensity of the focal-loss significantly influenced the accuracy-sensitivity trade-off: the aggressive scheme yielded a high macro-AUC (0.871) with lower recall for the majority class, while the moderated scheme resulted in a considerably more balanced 76.4% accuracy and 0.961 macro-AUC. Grad-CAM provided evidence of lesion-centred attention for all classes. In the future, the individual contribution of the metadata branch will be isolated, segmentation preprocessing will be added, and generalisation will be validated using additional corpora like ISIC 2019/ISIC 2020 and cross-validation.

REFERENCES

[1] P. Tschandl et al., 'The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,' Scientific Data, vol. 5, 2018.
[2] A. Rezvantalab et al., 'Dermatologist level dermoscopy skin cancer classification using different deep learning CNN algorithms,' arXiv:1810.10348, 2018.
[3] N. Gessert et al., 'Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data,' MethodsX, vol. 7, 2020.
[4] N. Gessert et al., 'Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting,' IEEE Trans. Biomed. Eng., vol. 67, no. 2, pp. 495-503, 2020.
[5] N. Ahmad et al., 'A novel framework of multiclass skin lesion recognition using deep learning and explainable AI,' Frontiers in Oncology, vol. 13, 2023.
[6] J. Frederich et al., 'Skin lesion classification using EfficientNet B0 and B1 via transfer learning for CAD,' AIP Conf. Proc., vol. 3080, 2024.
[7] C. Kavitha et al., 'Skin cancer detection and classification using deep learning techniques,' Procedia Computer Science, vol. 235, pp. 2793-2802, 2024.
[8] E. Gayatri, S. L. Aarthy, 'Reduction of overfitting on the highly imbalanced ISIC-2019 skin dataset using deep learning,' J. X-Ray Sci. Technol., 2024.
[9] G. H. Dagnaw et al., 'Skin cancer classification using vision transformers and explainable AI,' J. Med. Artif. Intell., vol. 7, 2024.
[10] S. Al-Fahdawi, M. I. Khalaf et al., 'A deep learning framework for early diagnosis and classification of skin cancer lesions in dermoscopy,' Scientific Reports, 2025.
[11] S. Cherukuri, S. D. Vara Prasad, 'EnsembleSkinNet: transfer learning-based skin cancer detection with explainable AI,' Frontiers in Oncology, 2025.
[12] M. A. Naeem et al., 'Automated skin cancer detection using MedFusionNet with attention-based fusion of ConvNeXt and ViT,' Scientific Reports, 2025.
[13] B. Kishore Kumar et al., 'Exploring skin lesion classification through explainable AI: Grad-CAM for interpretability,' Proc. ICACECS 2024, Springer SIST vol. 117, 2026.
[14] N. Ashfaq et al., 'SkinSight: Advancing deep learning for skin cancer diagnosis and classification,' Discover Computing, vol. 28, 2025.
[15] M. A. Khan et al., 'An EfficientNetB4-based CNN model with Soft Attention for skin lesion classification,' Eng. Technol. Appl. Sci. Res., 2026.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Lightweight Representative Grid Framework for Sparse Recommendation Systems

Sakshi Siva Ramakrishna*, T Anuradha

**Dept. Of CSE, Acharya Nagarjuna University, ramakrishna_sakshi@yahoo.co.in*

Dept. Of CSE, RVR&JC College of Engineering, a_talasila@yahoo.com

Abstract.

Recommendation systems rely on user-item interaction data that is usually sparse. There are numerous methods to deal with sparsity, but with limitations. Matrix factorization is a widely used method. It demands high computations and is limited by low interpretability. This work introduces a low-complexity approach that preserves structural information while avoiding latent factor optimization. This approach first identifies a grid band that captures significant user and item interaction patterns and uses it as a basis to make predictions. Experiments were conducted on the MovieLens 100K benchmark dataset using five-fold cross-validation. The results demonstrate competitive prediction performance and top-*K* recommendation quality compared with existing approaches. The proposed approach offers a computationally efficient alternative for recommendation tasks in sparse data situations.

Keywords. Representative Grid Recommender (RGR), Matrix Factorization, Structural Recommendation, Top-*K* Recommendation, Sparse Rating Prediction.

1. INTRODUCTION

Sparse matrices are data sources for recommendation systems, with much of the data remaining unobserved. The primary goal of a recommendation algorithm is to process this data to derive useful recommendations. Matrix Factorization (MF) has emerged as one of the most influential approaches owing to its ability to estimate missing ratings using a latent feature space **[1]– [3]**. However, sparsity limits accuracy, and learning through latent factors limits interpretation. Moreover, MF performs iterative parameter tuning, with high computational complexity for large-scale recommendation tasks **[1], [4]**.

This paper presents a lightweight recommendation approach for sparse rating prediction that uses a grid structure to pick representative users and items to guide the estimations. Prediction performance and recommendation quality are assessed using appropriate metrics.

The major contributions of this work are summarized as follows.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. A novel Representative Grid Recommender (RGR) is proposed instead of latent factor optimization.
2. A three-case prediction strategy is developed to improve recommendation robustness under highly sparse rating conditions.
3. The proposed framework preserves the structural characteristics of the observed rating matrix during the prediction process.
4. An experimental setup is presented to demonstrate the effectiveness, stability, and practical applicability of the proposed model.

2. RELATED WORK

Existing recommendation approaches can be classified into 4 categories.

2.1. Neighbourhood-based collaborative filtering

These approaches exploit similarities among users or items. They study similar users or relationships among similar items. Recommendation neighbourhoods are constructed using similarity metrics [5]– [8]. However, these techniques often exhibit reduced performance on large-scale sparse datasets where the neighbourhoods are also sparse [6], [9].

2.2. Matrix Factorisation recommender

Matrix Factorization constructs a product of low-dimensional latent user and item matrices to estimate unknown ratings [1], [3]. Numerous variants have evolved, ranging from biased Matrix Factorization to Neural Collaborative Filtering [4], [10]– [12]. However, these approaches are computationally expensive and less interpretable, particularly when sparseness is too high.

2.3. Hybrid and deep models

Hybrid frameworks combine collaborative filtering with content information, knowledge graphs, graph neural networks, transformer architectures, and large language models [13]–[15]. The hybridisation benefited from the positive elements of the constituent models. However, large computational resources, complex training pipelines, and extensive parameter optimization remain their major limitations.

2.4. Sparse rating prediction

Matrix completion, graph-guided recommendation, trust-aware recommendation, and low-rank approximation belong to this category [11], [12],[15]. However, these approaches generally rely on iterative parameter optimization, demanding significant computational resources.

All these observations motivate the design of better strategies to mitigate the problems associated with the existing methods. The proposed design is constrained by low computational complexity and better interpretability.

3. PROPOSED REPRESENTATION GRID RECOMMENDER

The observed rating matrix is of the form $R = [r_{ui}]_{m\times n}$, (1) ,

where

$$r_{ui} = \begin{cases} 1, \dots, k, & \text{if user } u \text{ rated item } i, in\ the\ range\ 1\ to\ k \\ 0, & \text{otherwise.} \end{cases}$$

The objective is to estimate the rating $\hat{r}_{ui}$. A grid is a subset of the rating matrix, with a minimal binary disjunction of rows forming a one-vector row band. Similarly, a minimal collection of columns forms a column band of the grid. This grid beam is a combination of representations of users (U_R) and items (I_R) that guide our prediction. These users and items are significant contributors to the user-item interaction behaviour.

Let $b_u$is a vector corresponding to user u. Starting from an all-zero vector,$W^{(0)} = 0$, successive logical disjunctions form

$$W^{(t)} = W^{(t-1)} \vee b_{u_t},\ t = 1,2, \dots, p, \quad (2)$$

The process continues until one vector is formed. The corresponding user set $U_R = \{u_1, u_2, \dots, u_p\}$ forms the representative user set.

Similarly, a dual procedure is applied to the binary item vectors to obtain the representative item set $I_R = \{i_1, i_2, \dots, i_q\}$. The Cartesian intersection of U_R and I_R constitutes the Representative Grid.

The complete prediction model is

$$\hat{r}_{ui} = \begin{cases} \hat{r}_{ui}^{(1)}, & u \in U_R \wedge\ i \notin I_R, \\ \hat{r}_{ui}^{(2)}, & u \notin U_R\ \wedge\ i \in I_R, \\ \hat{r}_{ui}^{(3)}, & \text{otherwise.} \end{cases} \quad (3)$$

If the target cell comes under the representative user set (U_R), the missing rating is estimated from representative neighbors $N_u = \{v_1, v_2, \dots, v_k\}$ who rated the target item. Therefore

$$\hat{r}_{ui}^{(1)} = \frac{1}{k} \sum_{v \in N_u} r_{vi}\,. \quad (4)$$

If the target cell comes under the representative item set (I_R), the missing rating is estimated from representative neighbors $N_i = \{j_1, j_2, \dots, j_l\}$. who were rated by the target user.

Therefore $\hat{r}_{ui}^{(2)} = \frac{1}{l} \sum_{j \in N_i} r_{uj}\,.$ (5)

If the target cell is out of the grid, the prediction falls under user-item statistics.

$$\hat{r}_{ui}^{(3)} = \frac{E_u + E_i}{2}. \quad (6)$$

, where E_u and E_i are average observed ratings of similar users and items. The similar users are those who rated other items similarly to the targeted user. Similar items also follow the similar criterion.

4. EVALUATION, RESULTS AND DISCUSSION

4.1. Prediction accuracy

The proposed Representative Grid Recommender (RGR) was evaluated against existing baselines using five-fold cross-validation on the MovieLens 100K dataset. Each fold has a train-to-test ratio of 80:20. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are the metrics for the evaluation. Table 1 presents the prediction accuracy results given by the proposed process in comparison with Matrix Factorization, user KNN, and item KNN models.

Table 1
Prediction Accuracy Comparison Between MF and RGR

Method	RMSE	MAE
MF	0.9376	0.7256
RGR	1.006267	0.806471
User KNN	1.005402	0.780696
Item KNN	0.98	0.899

It can be noticed that RGR achieves a matching prediction error with existing methods. The closeness of RMSE and MAE indicates that the proposed structural prediction mechanism provides a significant estimate of missing ratings under sparse conditions using a non-iterative, low-cost process. Unlike MF, RGR directly uses the observable interaction structures in the rating matrix.

4.2. Ranking performance

In recommendation systems, the quality of recommendations is also a significant contribution of a recommendation model. Table 2 presents the ranking performance comparison.

Table 2

Top-10 Ranking Performance Comparison on MovieLens 100K

Method	Precision@10	Recall@10	Hit Ratio@10	MAP@10	NDCG@10
User-KNN	0.019822	0.026241	0.167996	0.009718	0.025243
Item-KNN	0.007501	0.007300	0.066316	0.003210	0.008927
MF	0.043858	0.033038	0.287108	0.020119	0.049430
RGR	0.016214	0.014526	0.133400	0.011232	0.024597

RGR reached a Precision@10 of 0.016214, Recall@10 of 0.014526, Hit Ratio@10 of 0.133400, MAP@10 of 0.011232, and NDCG@10 of 0.024597. Although these values are lower than those of the MF model, the proposed method consistently outperformed Item-KNN in all ranking metrics. Compared with User-KNN, RGR got a higher MAP value, indicating improved ranking quality. The NDCG value of RGR (0.024597) is also comparable to that of User-KNN (0.025243), demonstrating that the proposed

representative-grid strategy preserves the relative ordering of relevant recommendations with a lighter processing framework.

It can be observed that the proposed method was evaluated for competitive recommendation performance. RGR showed better recommendation metric values in some sub-cases and was competitive in others. Without high-level computations or latent factors, the proposed approach can provide reliable recommendations. Therefore, RGR is a practical and interpretable recommendation framework.

Recommendation quality can be further improved by adopting structural extraction methods like Hyperlink Induced Topic Search (HITS).

5. CONCLUSION

As a structural representation model, RGR provides interpretable knowledge for recommendation without learning latent factors. By identifying users and items that contribute a significant basis for interaction behavior, this grid-based approach made predictions without iterative computations, reducing complexity. The proposed method is a lightweight alternative to complex matrix imputation methods. As with other existing methods, the proposed method has limitations when the data is too sparse. This limitation can be mitigated by adopting lightweight structural methods like HITS. The idea introduced in this paper motivates further lightweight approaches that reduce the burden on computing resources and improve interpretability and reliability.

6. REFERENCES

[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009, doi: 10.1109/MC.2009.263.

[2] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," in *Proc. IEEE Int. Conf. Data Mining*, 2008, pp. 263–272, doi: 10.1109/ICDM.2008.22.

[3] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback," in *Proc. UAI*, 2009, pp. 452–461.

[4] R. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 3rd ed. Cham, Switzerland: Springer, 2022, doi: 10.1007/978-1-0716-2197-4.

[5] P. Resnick *et al.*, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," in *Proc. ACM CSCW*, 1994, pp. 175–186, doi: 10.1145/192844.192905.

[6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," in *Proc. WWW*, 2001, pp. 285–295, doi: 10.1145/371920.372071.

[7] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005, doi: 10.1109/TKDE.2005.99.

[8] C. C. Aggarwal, *Recommender Systems: The Textbook*. Springer, 2016, doi: 10.1007/978-3-319-29659-3.

[9] O. A. S. Ibrahim, E. M. G. Younis, E. A. Mohamed, and W. N. Ismail, "Revisiting Recommender Systems: An Investigative Survey," *Neural Computing and Applications*, vol. 37, pp. 2145–2173, 2025, doi: 10.1007/s00521-024-10828-5.

[10] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural Collaborative Filtering," in *Proc. WWW*, 2017, pp. 173–182, doi: 10.1145/3038912.3052569.
[11] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization," in *Advances in Neural Information Processing Systems*, 2007.
[12] S. Sedhain, A. K. Menon, S. Sanner, and L. Xie, "AutoRec: Autoencoders Meet Collaborative Filtering," in *Proc. WWW Companion*, 2015, pp. 111–112.
[13] W. Fan *et al.*, "Graph Neural Networks for Social Recommendation," in *Proc. The Web Conference (WWW)*, 2019, pp. 417–426.
[14] J. Lin *et al.*, "How Can Recommender Systems Benefit from Large Language Models: A Survey," *ACM Transactions on Information Systems*, vol. 43, no. 2, 2025, doi: 10.1145/3678004.
[15] R. Mehta, S. Mishra, and S. Saha, "Non-Negative Matrix Factorization in Recommender Models: Concept, Survey, and Future Direction," *IEEE Access*, vol. 13, pp. 192492–192517, 2025, doi: 10.1109/ACCESS.2025.3630750.

Biographies



Sakshi Siva Ramakrishna received the MCA degree from IGNOU in 2002. Received M.Tech (CSE) from JNTU. Currently, he is a research scholar in the Department of Computer Science and Engineering, ACHARYA NAGARJUNA UNIVERSITY. His research areas include Data science and recommender systems.



T Anuradha received the MS degree in Software Systems from BITS, Pilani in 1997, MTech in CSE from Acharya Nagarjuna University in 2010, and the PhD degree in Cryptography from Sri Padmavathi Mahila Visva Vidyalayam, Tirupati in 2015. She is currently working as an Associate Professor at the Department of Computer Science and Engineering, Faculty of Engineering, at RVR&JC College of Engineering. Her research areas include cryptography and network security.

IC-DAIS
<Conference Series name> <Volume number and Year> <DOI Number>

Artificial Intelligence-Enabled Organizational Resilience and Corporate Social Responsibility: A Quantitative Study of Sustainability in Moroccan Small and Medium Enterprises in the Post-Pandemic Era

Iman El Joubari[1] , Fatima Ezzahra Khtabe[2]

[1]*Research Team in Territorial and Organizational Economics Faculty of Legal, Economic and Social Sciences, Tangier Abdelmalek Essaadi University, Morocco,* *eljoubariimane@gmail.com*
[2]*Independent researcher, MBA Candidate, Coventry University, United Kingdom.* *Khtabef@uni.Coventry.ac.uk*

Abstract.

Due to the increasing complexity of the business environment after the pandemic outbreak, the need for organizational resilience and sustainable business practices in small and medium enterprises (SMEs) has been heightened. This paper explores the impact of organizational resilience and Corporate Social Responsibility (CSR) on organizational sustainability in Moroccan SMEs and the potential moderating effect of Artificial Intelligence (AI) capabilities. A quantitative methodological design using questionnaire survey of 287 Moroccan SMEs has been used to validate the research model using Structural Equation Modeling (SEM). Both organizational resilience and CSR were found to have a positive impact on organizational sustainability. Moreover, AI capabilities positively moderate these impacts by improving the adaptive capacity, data-driven decision-making processes, and operational efficiency in order to respond effectively to changes in the business environment. Contribution to the literature is made through incorporation of AI, organizational resilience, and CSR in one comprehensive framework of organizational sustainability. The findings contribute to the existing literature in terms of practical implications related to digital transformation and organizational sustainability in an emerging economy.

Keywords. Organizational Resilience, Corporate Social Responsibility, Artificial Intelligence, SME Sustainability, Dynamic Capabilities.

1. INTRODUCTION

The current business world has become highly volatile due to global threats, technological changes, and the evolution of stakeholders' expectations. The emergence of such factors as the COVID-19 pandemic, geopolitical tensions, and economic instabilities has demonstrated numerous flaws in the operations of enterprises. As a result, companies have been forced to increase their resilience and introduce sustainable business models [1,4]. Thus, organizational resilience is a strategic capability that allows firms to predict any threats and

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

uncertainties, be ready to respond and recover while sustaining their operations and competitiveness [4,5].

Small and medium enterprises (SMEs) have a great significance for the economy since they are engaged in the generation of jobs, innovation activities, and economic growth. In particular, in Morocco, SMEs constitute around 95 percent of all registered enterprises and form the basis of the national economy. At the same time, these organizations experience many difficulties related to finances, technology, and institutions that hinder their ability to cope with various external shocks and achieve sustainable growth [2,3]. Thus, the enhancement of organizational resilience is vital for the sustainability and competitiveness of Moroccan SMEs. In turn, Corporate Social Responsibility (CSR) has become a strategic business approach that aims to develop stakeholder trust, corporate reputation, and organizational performance. According to the existing literature, CSR integration into the organizational strategy is able to positively affect the relations with stakeholders, responsible resource use, and sustainable business development in the case of SMEs functioning in developing countries [2,9]. Thus, CSR can be regarded as the mechanism that helps organizations balance their economic performance with environmental and social responsibilities.

The fast advancement of Artificial Intelligence (AI) has also affected the decision-making processes in organizations and their operation. Due to the implementation of AI technologies, including predictive analytics, intelligent automation, and decision support systems based on data, organizations can increase their effectiveness, predict threats, optimize the allocation of resources, and increase their adaptability [6,7]. For SMEs, the adoption of AI provides a great opportunity to increase their resilience due to the enhanced ability to make decisions, increase strategic flexibility, and innovate under dynamic conditions [6,10].

Even though the separate effects of organizational resilience, CSR, and AI capabilities have been researched separately in the literature, the effects of their integration on organizational sustainability remain understudied. Particularly, few empirical studies have been conducted for emerging economies, and even fewer studies have considered the role of AI capabilities in the relationships between organizational resilience, CSR, and sustainability [2,4].

Therefore, in order to fill in the knowledge gap, the current study explores the direct impact of organizational resilience and CSR on the sustainability of Moroccan SMEs by considering the effect of AI capabilities as a moderator using Structural Equation Modeling (SEM). The integration of the concepts of organizational resilience, CSR, and AI in the unified framework will allow the expansion of the research on sustainable business management and digital transformation.

2. METHODOLOGY

Research Design and Sampling Procedure

The present study employed the quantitative and cross-sectional survey approach in order to investigate the impact of organizational resilience, corporate social responsibility, artificial intelligence and organizational sustainability on Moroccan small- and medium-sized

enterprises (SMEs).The structured questionnaire was distributed from January to March 2026 through a hypothetico-deductive research design approach. The population of the study consisted of all SMEs listed with the Moroccan General Confederation of Enterprises (CGEM) and the Maroc PME Development Agency. Stratified random sampling was used to sample respondents from different industries and geographic locations. Number of valid responses was 287 with response rate 34.2%. Demographic details of the respondents are provided in Table 1.

Measurement Instruments

Constructs used in the study were quantified based on reliable and well-established measures. The resilience of the organization was measured from cognitive, behavioral, and contextual perspectives based on scales adopted from Koporcic et al. (2026) and Pasha et al. (2025) [8]. Measures of CSR practices were adopted from frameworks presented by Gnaoui and Moutahaddib (2024) and Vázquez et al. (2013) [9]. AI capabilities and sustainability of organizations were quantified using multidimensional five-point Likert-type scale adopted from the existing literature [6,10]. In order to establish the content validity and clarity of the questionnaire, it was pilot tested among 30 managers of SMEs.

2.1. About Data Analysis

Analysis was done through many processes including detailed statistics, assessing the reliability with Cronbach's alpha, correlation analysis, Structural Equation Modeling, confirmatory factor analysis , and structural equation modeling (SEM) was analyzed with maximum likelihood estimation. Analysis was conducted through statistical tools such as SPSS 28.0 and AMOS 28.0. The adequacy of models is checked through fit indices like χ^2/df, CFI, TLI, and RMSEA with values > 0.90 for CFI and TLI and < 0.08 for RMSEA.

3. RESULTS

3.1. Descriptive Statistics and Reliability Analysis

The detailed statistics and reliability analysis of the review constructs have been displayed in Table below. As can be seen from the results presented above, there is a moderate-to-high organizational resilience level amongst Moroccan small and medium-sized enterprises with average scores ranging from 3.42 (Contextual Resilience) to 3.68 (Cognitive Resilience). Average score of CSR practices was 3.51 (SD=0.74), while AI capabilities had an average score of 3.23 (SD=0.89), thus indicating a moderate level of AI capabilities among the sampled SMEs. Moreover, all values of Cronbach's alpha coefficient were higher than the required value of 0.70.

3.2. Confirmatory Factor Analysis

There was a good fit of the measurement model in the data: $\chi 2/df = 2.14$, CFI = 0.94, TLI = 0.93, RMSEA = 0.063 (90% CI: 0.052-0.074). There was also statistical significance of all factor loadings, which were larger than 0.60 and were equal to $p < 0.001$. It proved the presence of convergent validity in the measurement model. The Average Variance Extracted (AVE) was > 0.50 & its range was between 0.52 and 0.68. Composite reliability was greater than 0.70 and its range was between 0.84 and 0.92.The presence of discriminant validity was

shown using Fornell-Larcker criterion. The value of the square root of AVE for every construct was larger than correlation with other constructs

3.3. Correlation Analysis

Pearson correlation coefficient among constructs in Table. It can be seen from the table above that SME sustainability has a significant and positive relationship with all dimensions of organizational resilience ($r = 0.42–0.58$; $p < 0.001$), CSR practices ($r = 0.63$; $p < 0.001$), and artificial intelligence (AI) capabilities ($r = 0.47$; $p < 0.001$). Multicollinearity was not identified since all VIF values were lower than 3.0, which is within the acceptable range. The maximum inter-construct correlation was noted between CSR practices and Cognitive Resilience ($r = 0.61$) and was below the 0.70 cut-off value.

Table 1. Correlation Matrix Among Study Variables

Variable	1	2	3	4	5	6
1. Cognitive RES	1.00					
2. Behavioral RES	0.58***	1.00				
3. Contextual RES	0.52***	0.61***	1.00			
4. CSR Practices	0.61***	0.54***	0.48***	1.00		
5. AI Capabilities	0.43***	0.38***	0.35***	0.49***	1.00	
6. SME Sustainability	0.58***	0.52***	0.42***	0.63***	0.47***	1.00

Note: N = 287; *** $p < 0.001$; RES = Resilience

4. DISCUSSION

Based on the results obtained in the present study, the contribution of organizational resilience, Corporate Social Responsibility (CSR), and Artificial Intelligence (AI) capabilities to the sustainability of SMEs in Morocco is empirically justified. Positive correlations between different components of organizational resilience (cognitive, behavioral, and contextual) and sustainability of SMEs confirm the ability of resilient organizations to predict future disruptions, adapt to changing business environment, and ensure sustained organizational performance in the post-pandemic context. Therefore, these findings reaffirm the strategic importance of resilience as a key capability of any organization.At the same time, the research confirms the significant contribution of socially responsible activities of organizations to their sustainable performance. In this regard, the findings imply that socially responsible actions of organizations can help improve the trust of stakeholders, organizational reputation, and competitive advantage. Thus, CSR should be considered as a strategic resource that helps SMEs to survive and grow even in difficult conditions in which they operate.

Moreover, the research finds out that AI capabilities have positive effects on organizational adaptability and decision-making. In this way, the use of AI contributes to enhanced

sustainability of organizations through improvement of the effects of organizational resilience and CSR.

For SME managers, this means that investments in building of organizational resilience, integration of CSR into strategic planning, and implementation of AI technologies will contribute to enhanced organizational adaptability and sustainability. For the part of policymakers and development agencies, complementary measures to support SMEs in terms of improving digital infrastructure and organizing training programs in using AI technologies can be suggested.

Still there many challenges and limitation of an existing research work which can become an interest area for researcher. In particular, the use of cross-sectional design precludes making any conclusions about causality of relations between variables under investigation. Moreover, due to the narrow focus on Moroccan SMEs, the results cannot be generalized. Also, AI capabilities were investigated in a broad sense but not separately from other components of digitalization of organizations.

5. CONCLUSION

The current study investigates the link between organizational resilience, Corporate Social Responsibility (CSR), Artificial Intelligence (AI) capabilities, and organizational sustainability in the context of Moroccan SMEs applying a quantitative methodology. The obtained empirical results reveal that the cognitive, behavioral, and contextual factors of organizational resilience along with CSR initiatives positively influence SME sustainability. In addition, AI capabilities improve these associations through increasing organizational adaptability, efficiency, and making data-based decisions for achieving sustainable performance of businesses in a turbulent environment.The present research contributes to the relevant literature on SME sustainability in an emerging market by developing the integrated theoretical framework, which connects organizational resilience, CSR, and AI capabilities in order to explain the sustainability of firms. On the other hand, the results of this study have important practical implications for managers of SMEs concerning the significance of building resilience capabilities, incorporating CSR initiatives into organization strategy, and implementing digital technologies in their business processes.

In general, this research proves that the implementation of the discussed constructs has an essential positive influence on the sustainability of Moroccan SMEs. The future researches should further investigate the proposed framework applying longitudinal and cross-country data. The particular attention should be paid to the analysis of the effects of AI technologies on organizational sustainability.

6. REFERENCES

[1] H. Zine, E. M. Lotfi, M. Mahrouf, et al., "Modeling the spread of COVID-19 pandemic in Morocco," arXiv preprint arXiv:2010.04115, 2020.

[2] I. Gnaoui and A. Moutahaddib, "Towards an Inclusive Approach to Corporate Social Responsibility (CSR) in Morocco: CGEM's Commitment," arXiv preprint arXiv:2408.11519, 2024.

[3] T. Ahammed, M. Asad, and K. Sakib, "Impact of Fourth Industrial Revolution (4IR) on Small and Medium Enterprises (SMEs) and Employment in Bangladesh: Opportunities and Challenges," arXiv preprint arXiv:2412.21106, 2024.
[4] N. Koporcic, P. K. Kukkamalla, S. Markovic, and M. Bagherzadeh, "Resilience of small and medium-sized enterprises in times of crisis: An umbrella review," Review of Managerial Science, vol. 20, no. 1, pp. 1–42, 2026.
[5] Y. Trousselle, "Organizational resilience and sensemaking: A Weickian perspective," Management Research Review, vol. 37, no. 8, pp. 725–740, 2014.
[6] O. Agbaakin, "Leveraging Artificial Intelligence as a Strategic Growth Catalyst for Small and Medium-sized Enterprises," arXiv preprint arXiv:2509.14532, 2025.
[7] A. A. Shakarchi and M. A. Nafzaoui, "Leveraging AI for Organizational Resilience: Insights from Literature," International Journal of Economics and Management Studies, vol. 12, no. 3, pp. 45–58, 2025.
[8] A. Pasha, K. Morris, J. Nayem, et al., "Measurement Instruments Assessing Organizational Resilience in Health Facilities: A Systematic Review of Psychometric Properties," OpenRxiv, doi: 10.1101/2025.06.11.25329057, 2025.
[9] R. Vázquez, A. Carrasco, and M. E. López-Pérez, "Small and medium-sized enterprises and Corporate Social Responsibility: A systematic review of the literature," Quality & Quantity, vol. 47, no. 6, pp. 3205–3228, 2013.
[10] A. Akhir and F. R. Rahman, "The Convergence of AI, MIS, and Predictive Analytics: A Multi-Sectoral Review of Organizational Resilience," Pacific Journal of Advanced Engineering, vol. 3, no. 1, pp. 112–128, 2025.

Biographies

Dr. Iman El Joubari is a researcher in strategic management. Her research focuses on organizational resilience, SME sustainability, and organizational transformation through artificial intelligence. Her work explores the development of innovative, responsible, and resilient organizational models.

Fatima Ezzahra Khtabe is a researcher whose work focuses on strategic management, sustainable development, digital transformation, and entrepreneurship in the Middle East and North Africa (MENA) region. She is currently pursuing a Master of Business Administration (MBA) at Coventry University, following a Bachelor’s degree in Business Management with a specialization in International Business from Cardiff Metropolitan University. Her research explores the interactions between businesses, public policies, and sustainable development, with the aim of contributing to evidence-based decision-making and strengthening international cooperation.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Explainable Artificial Intelligence Framework for Smart Traffic Management and Emergency Response on National and State Highways

Dr. Archana J R, Dr. Suma S, Prof. Prathibha V, Dr Anjali K V, Mr. Ziyad Mohammed

Department of MBA, RNS Institute of Technology, Bangalore, India Department of MBA, Bangalore Integrated Management Academy (BIMA), Bangalore, India

Abstract.

Efficient traffic management and quick response actions to emergencies are crucial in order to provide safety on the roads and reduce congestion on national and state highways. Conventional traffic management methods tend to be rule-based, inflexible and unable to take decisions in real time. In this work, a novel traffic monitoring and emergency response solution by Explainable Artificial Intelligence (XAI) is proposed, by combining three main components: vehicle detection, congestion prediction, and incident identification and emergency route optimization. The plan includes machine learning and deep learning models with explainability methods to ensure that traffic authorities and emergency responders can make decisions with transparency and interpretability. Data from traffic sensors, CCTV cameras, GPS tagged vehicles, and weather forecasts are used to detect traffic anomalies and to forecast traffic congestion.

Keywords: Explainable Artificial Intelligence (XAI), Smart Traffic Management, Emergency Response, Intelligent Transportation Systems, Traffic Prediction, National and State Highways

1. INTRODUCTION

Urbanization, growing number of vehicles and the expansion of road networks have led to high rates of traffic jams and road accidents. Although the highways play vital roles in economic development and transportation of people and goods, the major issues are highway congestion, delayed emergency response time, inefficient incident management, fuel consumption, pollution and economic losses. The traditional traffic control schemes primarily rely on rigid and manually enforced regulations, which are insufficient for traffic flow and crisis situations. Hence, Intelligent Transportation Systems (ITS) must be able to monitor, predict and support decision-making to improve highway safety and efficiency in real-time [1], [2]. Artificial Intelligence (AI), machine learning, deep learning and computer vision are important technologies in ITS. Using information from CCTV cameras, traffic

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

sensors, GPS, connected cars, and weather stations, these techniques can be used for vehicle detection, prediction of traffic flows, analysis of traffic congestion, accident detection, and prioritization of emergency vehicles. In transportation applications, however, safety is a critical concern and the decisions of many AI models are hard to interpret and validate, as they are considered black-box systems [3, 4]. This transparency can enhance trust, help validate decisions, and help transportation agencies, emergency responders, and policy makers make informed decisions [5,6]. Quick response is essential to minimize death, injury and property damage after highway accidents. Ambulances, fire services and police vehicles can suffer delays due to congestion, lack of situational awareness and inefficient route planning. The integration of AI and IoT can facilitate real-time traffic monitoring, congestion forecasting, and adaptive emergency traffic management. The incorporation of XAI may help clarify route choice and prioritize emergency situations, further enhancing the reliability of operations and stakeholders' confidence. Unified platform for the interwoven integration of traffic monitoring, congestion prediction, incident detection, emergency response management, and explainable decision support. The plan is to enhance traffic efficiency, decrease the emergency response time and enable clear, reliable and trusted transportation decisions [9] [10].

2. LITERATURE REVIEW

2.1 Explainable Artificial Intelligence in Transportation

Artificial Intelligence (AI) is increasingly leveraging on Intelligent Transportation Systems (ITS) for traffic prediction, congestion monitoring and accident detection. But many of the AI models are black boxes, and their decisions are hard for transportation authorities to comprehend. Explicable Artificial Intelligence (XAI) is addressing this limitation by explaining the predictions of the AI. Molnar [11] noted that explainability is key to trust, accountability and regulatory compliance. Lundberg and Lee [12] presented an explanation-based approach called SHAP (SHapley Additive exPlanations), which measures the contribution of individual features to model predictions. These methods make machine learning and deep learning models more interpretable in transportation, among other uses.

2.2 AI-Based Traffic Prediction and Management

Over the last few years, there have been a number of research efforts to improve the monitoring and prediction of traffic with the help of deep learning, computer vision, reinforcement learning and sensor-based technologies. For real-time traffic management and anomaly detection, studies on traffic classification and prediction using explainable AI have highlighted the need for accurate predictions and interpretability [13]. Haydari and Yılmaz [16] surveyed deep learning and reinforcement learning for ITS, and presented their applications in adaptive traffic signal control, route optimization and autonomous decision-making. Nellore and Hancke [18] emphasized the need for wireless sensor networks (WSN) and the Internet of Things (IoT) to support real-time traffic monitoring and analyzing traffic congestion.

2.3 Explainable AI for Emergency Response and Autonomous Transportation

When it comes to emergency transportation applications, there's often a tight deadline to meet and dependable, accurate decision-making is imperative, otherwise the consequences could be dire. During emergencies, XAI can offer explanations for predictions and recommendations, enabling that decision to be validated by XAI. Degas et al. [15] explored AI and XAI in AIR and illustrated how explainable decision-making in AIR can be crucial for operational safety and reliability. Likewise, Kuznietsov et al. [17] highlighted the importance of applying XAI in the field of autonomous driving, where AI decision-making directly impacts on human lives. Grad-CAM [14] is a new method that generates visual explanations by highlighting areas of an image that are important for CNN predictions. These methods aid in the clear identification of accidents and traffic monitoring.

2.4 Problems and Future Directions

The use of AI, IoT, and real-time transportation data has been the subject of numerous recent studies. A multi-agent machine learning approach to the identification of accidents and emergency coordination was proposed by Ayesha et al. [19] and showed encouraging results. Alnowaiser et al. [20] proposed a GNN-LSTM model along with SHAP explainability for smart city traffic accident severity prediction. These methods can predict accurately and make predictions understandable at an interpretive level, but most research is based on urban contexts or specific modes of transportation.

3. RESEARCH METHODOLOGY

The proposed research is an Explainable Artificial Intelligence (XAI) Framework for Smart Traffic Management and Emergency Response on National and State Highways to enhance the traffic monitoring and traffic congestion prediction, incident detection and emergency vehicle routing.

3.1 Data Acquisition

The proposed architecture gathers heterogeneous traffic information from various sources such as CCTV cameras, traffic sensor networks, GPS-equipped vehicles, roadside IoT devices, weather forecasting stations, and databases from emergency service providers. Data collected comprises vehicle count, speed, traffic density, weather, location and time. The traffic data is encoded in the form of

$$D = \{x_1, x_2, x_3, \dots, x_n\} \qquad (3.1)$$

The data set D consists of a traffic record x_i and the number of observations is n. They are each stored as:

$$x_i = (v, s, d, w, l, t) \qquad (3.2)$$

v is number of vehicles, s is average speed, d is traffic density, w is weather, l is location, and t is the timestamp.

3.2 Data Preprocessing

Data is cleaned for missing data, duplicated data, noise and inconsistencies. If there are missing values they are replaced with the meaning:

$$x' = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad (3.3)$$

The numeric attributes are normalized with Min-Max normalization:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (3.4)$$

This normalizes the values from 0 to 1 and enhances the training of the model.

3.3 Feature Engineering

The traffic data includes vehicle counts, lane occupancy, average speed, weather conditions, accident records, road type and congestion levels. Redundant and highly correlated features are eliminated to enhance the prediction efficiency. The feature set selected is displayed as:

$$F = \{f_1, f_2, f_3, \dots, f_k\} \qquad (3.5)$$

where F is the selected features and k is the number of all features.

3.4 AI-Based Traffic Prediction Model

The processed data is fed into an AI model to predict congestion, detect abnormal traffic, and identify incidents. The forecast is given as:

$$\hat{Y} = f(X, \theta) \qquad (3.6)$$

where x is the input features and θ is the model parameters. Congestion probability is calculated using:

$$P = \frac{1}{1 + e^{-z}} \qquad (3.7)$$

where $z = w^T X + b$. If $P > 0.5$, the traffic condition is classified as congested; otherwise, it is considered normal.

4. Results and Analysis

The proposed explainable artificial intelligence (XAI) framework was evaluated for the effectiveness for smart traffic management and emergency response on national and state highways. The performance measures used were the standard classification performance

measures (Accuracy, Precision, Recall, F1-Score) and operational performance measures (Congestion detection rate, emergency response time).

Table 2. Performance Comparison of Traffic Prediction Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Decision Tree	90.24	89.10	88.45	88.77
Random Forest	93.16	92.85	92.10	92.47
Support Vector Machine	94.08	93.71	93.28	93.49
XGBoost	95.64	95.20	94.86	95.03
Proposed XAI Framework	**97.82**	**97.35**	**97.01**	**97.18**



Figure 2: Performance Comparison of Traffic Prediction Models

The following tables and figure will compare the proposed Explainable AI framework with the other machine learning models for traffic prediction. The proposed framework was found to have the highest Accuracy (97.82%), Precision (97.35%), Recall (97.01%), and F1-Score (97.18%) which indicates that the framework has a good ability to detect traffic congestion and incidents.

Table 3. Operational Performance of the Proposed Smart Traffic Management Framework

Performance Metric	**Conventional System**	**Proposed XAI Framework**
Congestion Detection Rate (%)	89.42	**97.21**
Incident Detection Rate (%)	90.15	**98.04**
Average Emergency Response Time (min)	12.8	**7.3**
Route Optimization Accuracy (%)	91.56	**98.18**
Decision Transparency Score (%)	65.30	**96.75**

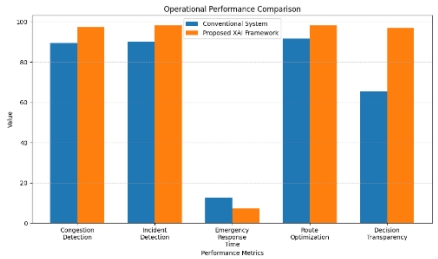


Figure 3: Operational Performance Comparison

The operational performance of the proposed framework is given in Table 3 and figure 3. The XAI-based system is able to greatly improve congestion mitigation, incident detection and reduce the average emergency response time from 12.8 minutes to 7.3 minutes, compared to the traditional traffic management system.

4. CONCLUSION

The purpose of this study is to create Explainable Artificial Intelligence (XAI) system to Smart Traffic Management & Emergency Response in National and State Highway. The framework aims to make real-time traffic monitoring, real-time congestion prediction using AI, real-time incident detection, providing explainable decision support for the traffic system, and real-time emergency route optimization, into a smart transport system.

5. REFERENCES

[1]. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 1135–1144. https://arxiv.org/abs/1602.04938

[2]. K. A. Sattar, I. Ishak, L. S. Affendey, and S. N. B. M. Rum, "Road crash injury severity prediction using a graph neural network framework," *IEEE Access*, vol. 12, pp. 37540–37556, 2024, doi: 10.1109/ACCESS.2024.3373885.

[3]. O. I. Aboulola, E. A. Alabdulqader, A. A. AlArfaj, S. Alsubai, and T.-H. Kim, "An automated approach for predicting road traffic accident severity using transformer learning and explainable AI technique," *IEEE Access*, vol. 12, pp. 61062–61072, 2024, doi: 10.1109/ACCESS.2024.3380895.

[4]. O. I. Aboulola, "Improving traffic accident severity prediction using MobileNet transfer learning model and SHAP XAI technique," *PLOS ONE*, vol. 19, no. 4, art. e0300640, 2024, doi: 10.1371/journal.pone.0300640.

[5]. J. Sunkpho, C. Se, W. Wipulanusat, and V. Ratanavaraha, "SHAP-based convolutional neural network modeling for intersection crash severity on Thailand's highways," *IATSS Research*, vol. 49, no. 1, pp. 27–41, 2025. https://doaj.org/article/6b42db1a168947079d29b4614958b6ec

[6]. A. Budzyński and A. Czerepicki, "Towards sustainable road safety: Feature-level interpretation of injury severity in Poland (2015–2024) using SHAP and XGBoost," *Sustainability*, vol. 17, no. 17, art. 8026, 2025. https://www.mdpi.com/2071-1050/17/17/8026

[7]. S. Dong, A. Khattak, I. Ullah, J. Zhou, and A. Hussain, "Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley additive

explanations," *Int. J. Environ. Res. Public Health*, vol. 19, no. 5, art. 2925, 2022, doi: 10.3390/ijerph19052925. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8910532/

[8]. T. Nguyen, A. Canossa, and J. Zhu, "How human-centered explainable AI interfaces are designed and evaluated: A systematic survey," *arXiv preprint*, arXiv:2403.14496, 2024. https://arxiv.org/abs/2403.14496

[9]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[10]. F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint*, arXiv:1702.08608, 2017. https://arxiv.org/abs/1702.08608

[11]. C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2019. https://christophm.github.io/interpretable-ml-book/

[12]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, 2017, pp. 4765–4774. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[13]. A survey on explainable artificial intelligence for internet traffic classification and prediction, and intrusion detection," *IEEE Communications Surveys & Tutorials*, 2024. https://ieeexplore.ieee.org/document/10763502/

[14]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[15]. A. Degas, M. R. Islam, C. Hurter, S. Barua, H. Rahman, M. Poudel, D. Ruscio, M. U. Ahmed, S. Begum, M. A. Rahman, *et al.*, "A survey on artificial intelligence (AI) and explainable AI in air traffic management: Current trends and development with future research trajectory," *Applied Sciences*, vol. 12, no. 3, p. 1295, 2022. https://www.mdpi.com/2076-3417/12/3/1295

[16]. A. Haydari and Y. Yılmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 11–32, Jan. 2022, doi: 10.1109/TITS.2020.3008612.

[17]. A. Kuznietsov, B. Gyevnar, C. Wang, S. Peters, and S. V. Albrecht, "Explainable AI for safe and trustworthy autonomous driving: A systematic review," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 12, pp. 19342–19364, Dec. 2024, doi: 10.1109/TITS.2024.3474469.

[18]. K. Nellore and G. P. Hancke, "A survey on urban traffic management system using wireless sensor networks," *Sensors*, vol. 16, no. 2, p. 157, 2016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4801535/

[19]. S. Ayesha, A. Aslam, M. H. Zaheer, and M. B. Khan, "CIRS: A multi-agent machine learning framework for real-time accident detection and emergency response," *Sensors*, vol. 25, no. 18, p. 5845, 2025. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12473270/

[20]. K. Alnowaiser, "A computational intelligence GNN–LSTM framework for spatiotemporal prediction of traffic accident severity in smart cities using SHAP XAI," *Int. J. Comput. Intell. Syst.*, vol. 18, art. 271, 2025, doi: 10.1007/s44196-025-01015-y.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Cloud-Native Data Fabric Architecture for Unified Risk and Fraud Intelligence in Banking Systems

Suresh Varma Dendukuri

Anna University, India, sureshvashishtad@gmail.com

Abstract.

Unified fraud intelligence has become essential in modern banking by enabling joint reasoning over transactional and relational financial data for applications such as fraud detection, Anti-Money Laundering (AML), and risk scoring. Current rule-based, statistical, graph-based and recent deep learning or Large Language Model (LLM) based methods usually work on financial data in silos, resulting in data fragmentation and poor cross-domain generalization. To overcome these problems, we introduce the concept of a unified cloud-native data fabric framework, namely, UniRisk-Fabric, which builds multi-source financial data in a unified way through common representations. The framework is comprised of two layers: (i) the Data Fabric Layer to achieve cross-source data integration and alignment to minimize data fragmentation; and (ii) the Cross-Domain Intelligence Layer to learn unified representations for enhanced transferability across the diverse financial domains. As seen from the experimental evaluation, UniRisk-Fabric is able to achieve 89% accuracy on the IBM AML dataset and 86% on the IEEE Fraud Detection dataset consistently outperforming traditional, graph-based and LLM-based methods in the unified fraud intelligence.

Keywords. Fraud, Financial, Intelligence, Large Language Model.

1. INTRODUCTION

Unified Fraud Intelligence allows to correlate different financial data types, such as transactional data and relational transaction networks to facilitate fraud detection, Anti-Money Laundering (AML), and financial risk analysis [1, 2]. In recent years, the use of graph neural networks (GNNs) and deep learning has made it easier to detect patterns of fraud, and more accurately—because it accounts for structural relationships and high-dimensional representations of features—[3]–[6]. Financial intelligence applications are further improved by large language models (LLMs) in the area of contextual reasoning and representation learning [7, 8]. However, current solutions mostly operate in a transactional/comprising relational style, causing data silos and lack of cross domain knowledge transfer. This means that the performance of models can suffer in diverse financial settings and changing fraud patterns [9]–[11]. To overcome these drawbacks, we will propose UniRisk-Fabric, a unified cloud-native data fabric architecture that allows to connect and access heterogeneous financial data sources through shared cross domain representations. It proposes two stages: (i) Data Fabric Layer where cross-source data integration and alignment are achieved to alleviate data fragmentation, and (ii) Cross-

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Domain Intelligence Layer which learns compact and transferable representations to enhance robust and generalization capabilities. Experimental results for both the IEEE Fraud Detection and the IBM AML Transactions datasets show that UniRisk-Fabric outperforms traditional machine learning and graph-based methods as well as LLM-based methods in terms of both performance and cross-domain generalization and computational efficiency.

The main contributions of this work are summarized as follows:

- We propose UniRisk-Fabric, a two stage unified data fabric framework where transactional and relational financial data are unified with shared representations to decrease data fragmentation and enhance cross-domain learning;
- The extensive experiments show that UniRisk-Fabric can consistently outperform the conventional machine learning, graph-based and LLM-based baselines on the IEEE Fraud Detection and IBM AML datasets, and that it is more robust, efficient in representation and can cross domain generalize;

2. RELATED WORKS

Financial fraud detection has shifted from basic machine learning to sophisticated graph-based and deep learning methods that can detect more complex fraudulent activities. Early Anti-Money Laundering (AML) and financial forensic studies showed the usefulness of graph representations to model the relationship between transactions and to identify suspicious financial activity [1]. More recent survey studies underscore the prevalence of using machine learning techniques for fraud detection, as well as challenges including data imbalance, changing fraud patterns and lack of cross domain generalization [2] [9] [10]. Despite the advances of traditional deep learning models such as variational autoencoders and others neural architectures that enhance the representation of the features, they are mainly applied to individual transactions without taking into account the relational information between them [11]. These constraints have led to the development of an alternative approach to learning called graph-based learning as a means of representing financial transaction networks. Dynamic graph learning, heterogeneous graph neural networks, evolving graph convolutional networks and graph attention mechanisms have shown more powerful potential for modeling financial structural relations and temporal dependence in financial systems [3]–[6], [12]–[15]. In recent years, there have also been developments in the use of Large Language Models (LLMs) to improve the aspects of contextual reasoning and representation learning for complex financial intelligence tasks [7] and [8]. Notwithstanding such progress, the existing methods mostly work with transactional and relational information separately, and data is fragmented and it is not possible to move knowledge between a range of heterogeneous financial environments. These constraints inspire the proposed UniRisk-Fabric, a novel data fabric architecture that is designed to unify multi-source financial data sources for enhanced cross-domains representation learning, robustness, and fraud intelligence.

3. METHODOLOGY

This section introduces the UniRisk-Fabric approach to tackle data fragmentation and cross-domain generalization in fraud intelligence (Fig. 1). The framework considers fraud detection as a single learning task from the non-homogeneous financial data, and uses features extracted from the transactions as well as features extracted from the relations graph to predict fraudulent behaviors. In contrast to the usual methods that work on these data sources separately, UniRisk-Fabric learns a common representation that allows for effective knowledge transfer between heterogeneous financial domains. Stage I (Data Fabric Layer) is the stage that cross-source integrates, normalizes, and aligns the data to build a single

representation of both transactional and relational data as shown in Fig. 1. The shared representation is further refined through masking in Stage II (Cross-Domain Intelligence Layer) which chooses the most informative parameters. The optimization ensures computational efficiency, predictive accuracy and sparsity, enhancing generalization and robustness to noise. Thanks to the resulting compact parameter subset, fraud intelligence will be accurate and transferable in the various financial contexts.

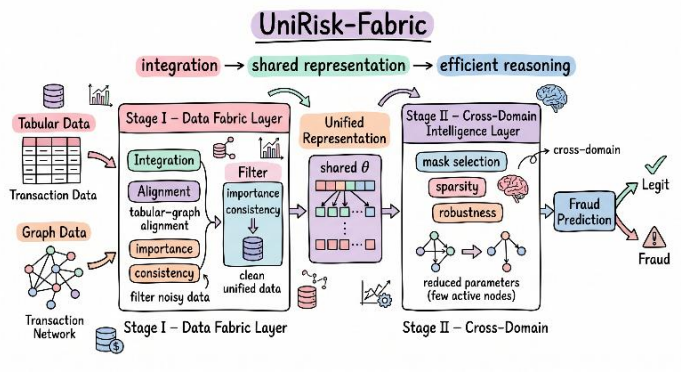


Figure 3.1. Overview of the UniRisk-Fabric framework illustrating unified data integration and cross-domain intelligence for robust and efficient fraud prediction.

The overall design of a data center can be classified in 4 categories Tier I-IV each one presenting advantages and disadvantages related to power consumption and availability. In most cases availability and safety issues yield to redundant N+1, N+2 or 2N data center designs and this has a serious effect on power consumption. According to Figure2.1, a data center has the following main units

4. EXPERIMENTAL ANALYSIS

4.1. Datasets

The datasets used for the evaluation of UniRisk-Fabric are two financial datasets widely used for financial fraud detection and IBM AML Transactions, respectively (Table I). Complementary financial domains: These datasets include financial transactional data with high dimensionality and transactional networks that are relational in nature, so that a complete evaluation of the unified fraud intelligence can be performed. The same data splits (70%/15%/15%) and preprocessing (as per previous studies) are used to ensure a fair comparison. The IEEE Fraud Detection dataset is an anonymized transaction, identity, and temporal attributes dataset for transaction-level fraud detection in the presence of class imbalance and feature noise, while the IBM AML Transactions dataset is a financial transaction network used to assess relational anti-money laundering analysis using graph-based representations of these networks. These datasets are used in conjunction to evaluate the proposed framework in the context of both transactional and graph-based financial settings.

4.2. Evaluation Metrics

Six complementary metrics are used to assess the performance of UniRisk-Fabric. Accuracy (Acc): It is the overall classification accuracy. Cross Domain Robustness (CDR): tests the model for robustness when it encounters data domain shifts between transactional and relational. The Cross-Domain Consistency (CDC) component quantifies the consistency of predictions when both the original inputs and the transformed versions are different. Cross-Domain Consistency (CDC) measures the stability of prediction across the discrepancies of

both the original and the transformed input, which captures the framework's ability to address data fragmentation. Fusion Efficiency (FE) is used to evaluate the compactness of the unified representation, and is computed as the parameter reduction, with low numbers signifying high efficiency. The measure of the accuracy retention (AR) reflects the level of fidelity in the compressed model's predictive accuracy compared to the full model. Finally, the Overall Score (OS) combines accuracy, robustness, consistency and efficiency in one score to give a balanced evaluation of the proposed framework.

4.3. Hyperparameters

The hyperparameters of UniRisk-Fabric are tuned to satisfy two criteria: predictive performance and robustness along with the computational efficiency (Table II). Importance and consistency thresholds are used in a data fabric layer (Stage I) to preserve features of high predictive sensitivity but low domain-to-domain variation across transactional and relational domains. During Stage II (Cross-Domain Intelligence Layer), sparsity and robustness regularization further enhance the unified representation with compactness and domain-invariant predictive ability in the face of domain shifts. The Adam optimizer, mini-batch learning and early stopping are used to optimize the masking variables. The Shared representation learning is done sequentially from Data Fabric Layer to Cross-Domain Intelligence Layer integrating tabular and graph-based features. To have the results of all the experiments comparable and reproducible, the same high-performance GPU is used throughout and the same batch sizes and training parameters are applied for all the baseline models.

5. RESULT ANALYSIS

5.1. Comparison with State-of-the-Art Methods

The above-mentioned datasets of IEEE Fraud Detection and IBM AML Transactions are used to compare UniRisk-Fabric with state-of-the-art methods in the evaluation metrics provided in Section IV (Table I). The traditional machine learning and graph-based approaches have been proven to struggle with cross domain generalization because they model only the transactional or relational data respectively. While these LLM-based models like LLaMA-2, Mistral, and Gemma are competitive in predictive tasks, their performance across diverse financial tasks and domains is restricted and needs enhancement. Contrary to this, UniRisk-Fabric consistently outperforms on Accuracy, Cross-Domain Robustness (CDR), Cross-Domain Consistency (CDC), and Accuracy Retention (AR) and decreases Fusion Efficiency (FE); Fisherman's accuracy is 10% higher than the average accuracy of the other algorithms. The findings indicate that the proposed shared representation is able to effectively combine transactional, relational information, and that it allows the creation of compact, powerful, and transferable fraud intelligence, outperforming traditional, graph-based and LLM-based methods.

TABLE I
PERFORMANCE COMPARISON ACROSS BASELINES AND UNIRISK-FABRIC

Model	Data	Acc ↑	CDR ↑	CDC ↑	FE ↓	AR ↑	OS ↑
Logistic Regression	IEEE	0.842	0.611	0.598	0.72	0.81	0.716
Random Forest	IEEE	0.887	0.642	0.631	0.69	0.84	0.738
GCN	IEEE	0.861	0.701	0.689	0.66	0.83	0.748
GAT	IEEE	0.873	0.724	0.712	0.64	0.85	0.761
LLaMA-2	IEEE	0.918	0.681	0.663	0.62	0.88	0.772
Mistral	IEEE	0.924	0.695	0.671	0.60	0.89	0.780
Gemma	IEEE	0.912	0.688	0.667	0.61	0.87	0.775

UniRisk-Fabric	IEEE	0.945	0.812	0.801	0.48	0.93	0.872
Logistic Regression	IBM AML	0.801	0.588	0.572	0.74	0.79	0.702
Random Forest	IBM AML	0.846	0.623	0.601	0.71	0.82	0.728
GCN	IBM AML	0.872	0.751	0.734	0.63	0.85	0.769
GAT	IBM AML	0.885	0.768	0.752	0.61	0.86	0.781
LLaMA-2	IBM AML	0.903	0.702	0.681	0.59	0.88	0.777
Mistral	IBM AML	0.909	0.715	0.694	0.57	0.89	0.785
Gemma	IBM AML	0.897	0.708	0.689	0.58	0.87	0.779
UniRisk-Fabric	IBM AML	0.932	0.829	0.814	0.46	0.92	0.865

5.2. Cross-Domain Analysis

The cross-domain transfer performance of these models is assessed by training them on one dataset and testing them on another, and evaluating them on their ability to transfer knowledge to different financial domains. All baseline methods suffer degradation in performance when there is a change in distribution type between transactional and relational data. Domain shifts have the greatest impact on the traditional machine learning models, whereas graph-based and LLM-based models exhibit moderate cross-domain transferability but are sensitive to domain shifts and exhibit poorer cross-domain robustness and consistency. By contrast, UniRisk-Fabric consistently shows better results in both directions of transfer, with higher Accuracy, Cross-Domain Robustness (CDR), Cross-Domain Consistency (CDC) and Accuracy Retention (AR), and lower Fusion Efficiency (FE). The results indicate that proposed shared representation is able to consolidate the transferable knowledge in heterogeneous financial environments, thus, resulting in stronger and more reliable fraud intelligence.

5.3. Computational Analysis

Table II shows the detailed comparison of the computational efficiency between baseline methods and UniRisk-Fabric. One important finding is the significant reduction in parameter usage with the application of the UniRisk-Fabric to LLM-based models. Results in significant reduction in computational cost with full-scale models compressed into unified representations that are much smaller. The improvements in efficiency do not only result in fewer floating-point operations, but also in less inference latency and less memory usage. As an example, models with UniRisk-Fabric can obtain substantial inference speedup (latency reduction of almost two times) and throughput growth. In addition, memory requirements are significantly lowered, allowing the deployment in environments with limited resources. Furthermore, the energy efficiency of the framework is illustrated by the reduction in energy consumption per sample, indicating the framework's ability to perform inference at high speeds and low energy costs in practical financial systems. More expressive graph-based approaches have more computational overhead because of the need to pass messages over large scale networks, and have not yet got a good option to integrate with transactional data.

TABLE II
COMPUTATIONAL EFFICIENCY COMPARISON ACROSS MODELS

Model	**Params (M) ↓**	**FLOPs (G) ↓**	**Inference Latency (ms) ↓**	**Throughput (samples/s) ↑**	**Memory (GB) ↓**	**Energy / Sample (J) ↓**

Logistic Regression	0.05	0.01	2.1	5200	0.2	0.03
Decision Tree	0.10	0.02	2.8	4800	0.3	0.04
Random Forest	1.20	0.15	6.5	2200	0.8	0.09
XGBoost	1.50	0.18	7.2	2000	0.9	0.10
GCN	3.80	1.20	18.5	820	2.1	0.28
GAT	4.50	1.45	22.3	700	2.4	0.34
LLaMA-2 (Full)	7000	180	120.0	120	28.0	2.40
Mistral (Full)	7000	165	110.0	135	26.0	2.20
Gemma (Full)	2000	95	85.0	180	16.0	1.60
LLaMA-2 + UniRisk-Fabric	850	42	58.0	260	9.5	0.95
Mistral + UniRisk-Fabric	820	38	52.0	290	8.8	0.88
Gemma + UniRisk-Fabric	420	21	41.0	340	6.2	0.62

6. ABLATION STUDY

Table III compares the performance of the two main components of the UniRisk-Fabric across the two datasets, IEEE Fraud Detection and IBM AML Transactions. LLaMA-2 has a moderate level of predictive performance, yet it has a weaker robustness and consistency across domains, possibly as a result of its fragmented representations. The enhancement of cross-source integration and robustness, consistency, and more effective parameter utilization are achieved through the inclusion of Stage I (Data Fabric Layer). Stage II (Cross-Domain Intelligence Layer) improves the performance using sparsity-aware and robustness-aware optimization for compact and transferable representation. The complete UniRisk-Fabric framework is the most accurate (Accuracy), Cross-Domain Robust (CDR), Cross-Domain Consistent (CDC), and the most accurate (Accuracy Retention) but has the lowest Fusion Efficiency (FE). These findings validate the complementary contribution of both stages to enhance cross-domain generalization, robustness and computational efficiency.

TABLE IIIABLATION STUDY OF UNIRISK-FABRIC

Model Variant	**Dataset**	**Acc ↑**	**CDR ↑**	**CDC ↑**	**FE ↓**	**AR ↑**	**OS ↑**
LLaMA-2 (Base)	IEEE	0.856	0.642	0.621	0.64	0.86	0.724
+ Stage I (No Consistency Constraint)	IEEE	0.874	0.701	0.672	0.59	0.88	0.763
+ Stage I (Full: Importance + Consistency)	IEEE	0.901	0.756	0.734	0.54	0.90	0.804
+ Stage II (No Sparsity Constraint)	IEEE	0.914	0.801	0.772	0.58	0.91	0.824
+ Stage II (No Robustness Objective)	IEEE	0.892	0.713	0.689	0.52	0.89	0.781
UniRisk-Fabric (Full Model)	IEEE	0.945	0.812	0.801	0.48	0.93	0.872

7. CONCLUSION AND FUTURE WORKS

In this work, we have proposed a unified data fabric framework, named UniRisk-Fabric, to solve data fragmentation and cross-domain generalization problems in a financial fraud intelligence system, which consists of two stages. The experimental results show promising enhancements in terms of accuracy, robustness, consistency and model efficiency, across the two IEEE Fraud Detection and IBM AML Transactions datasets. The proposed framework has been tested on benchmark datasets and needs to be evaluated in large-scale

operational banking environments with changing fraud patterns and distributions of data across institutions and operational constraints. Additionally, data privacy and regulatory compliance, computational resources, and integration with existing banking systems are critical considerations for practical deployment. Going forward, future work will focus on extending framework to multimodal financial data, adaptive real-time data fabric architectures, and models with higher interpretability to further improve scalability, transparency and practical applicability.

8. REFERENCES

[1] M. Weber, G. Domeniconi, J. Chen, D. K. I. Weidele, C. Bellei, T. Robinson, and C. E. Leiserson, 'Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics', arXiv Preprint, 2019.

[2] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, 'Financial fraud detection through the application of machine learning techniques: A literature review', Humanities and Social Sciences Communications, vol. 11, no. 1, 2024.

[3] D. Cheng, Y. Zou, S. Xiang, and C. Jiang, 'Graph neural networks for financial fraud detection: A review', Frontiers of Computer Science, vol. 19, no. 9, 2025.

[4] [4] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, and C. Leiserson, 'EvolveGCN: Evolving graph convolutional networks for dynamic graphs', Proc. AAAI Conf. Artificial Intelligence, vol. 34, no. 4, pp. 5363–5370, 2020.

[5] S. M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart, 'Representation learning for dynamic graphs: A survey', Journal of Machine Learning Research, vol. 21, no. 70, pp. 1–73, 2020.

[6] Z. Liu, C. Chen, X. Yang, J. Zhou, X. Li, and L. Song, 'Heterogeneous graph neural networks for malicious account detection', Proc. ACM Int. Conf. Information and Knowledge Management (CIKM), pp. 2077–2085, 2018.

[7] S. Tripathi, M. T. Nafis, I. Hussain, and J. Gao, 'The confidence paradox: Can LLM know when it's wrong?', Proc. Int. Joint Conf. Natural Language Processing (IJCNLP), pp. 2078–2087, 2025.

[8] S. Tripathi, M. T. Nafis, I. Hussain, and A. K. J. Saudagar, 'Multimodal fine-tuning of LLMs for robust document visual question answering', IEEE Access, vol. 13, pp. 174611–174623, 2025.

[9] Y. Lucas and J. Jurgovsky, 'Credit card fraud detection using machine learning: A survey', arXiv Preprint, 2020.

[10] Z. S. Rubaidi, B. B. Ammar, and M. B. Aouicha, 'Fraud detection using large-scale imbalance dataset', International Journal on Artificial Intelligence Tools, vol. 31, no. 8, 2022.

[11] Y. Ding, W. Kang, J. Feng, B. Peng, and A. Yang, 'Credit card fraud detection based on improved variational autoencoder generative adversarial network', IEEE Access, vol. 11, pp. 83680–83691, 2023.

[12] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, 'Enhancing graph neural network-based fraud detectors against camouflaged fraudsters', Proc. ACM Int. Conf. Information and Knowledge Management (CIKM), pp. 315–324, 2020.

[13] I. Alarab, S. Prakoonwit, and M. I. Nacer, 'Competence of graph convolutional networks for anti-money laundering in Bitcoin blockchain', Proc. Int. Conf. Machine Learning Technologies, pp. 23–27, 2020.

[14] N. Innan et al., 'Financial fraud detection using quantum graph neural networks', Quantum Machine Intelligence, vol. 6, no. 1, 2024.

[15] A. G. Vrahatis, K. Lazaros, and S. Kotsiantis, 'Graph attention networks: A comprehensive review of methods and applications', Future Internet, vol. 16, no. 9, 2024.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

"Ensemble Knowledge Distillation for Federated Semi-Supervised Medical Image Classification"

Shivani S, Suma L

Dept. of CSE, Bangalore Institute of Technology, Bangalore 560002

shivani7632@gmail.com, sumal@bit-bangalore.edu.in

Abstract.

Federated Learning (FL) allows clients (e.g., hospitals) to train shared models while preserving privacy and not sharing raw data. Most FL methods assume all the data is labeled, which is not the case for medical imaging because of the high cost and the specialized skills needed to perform data annotation. Medical FL clients possess a large volume of unlabeled data but only a small amount of labeled data. In this work, we present EKDFSSL (Ensemble Knowledge Distillation with Federated Semi-Supervised Learning) that extends FSSL to the case of detecting pneumonia in X-ray images. Client models in FSSL gain the ability to learn from unlabeled data due to the presence of the global model as a teacher and the application of consistency regularization to the client models. This enhances the stability and reliability of the model predictions in the presence of input data variations. At the server side, knowledge distillation is performed to reduce overfitting and to improve model performance by aggregating the client models. This technique of Federated Learning extends the client models ability of aligning their predictions with the global model, improving generalization. Our method is privacy preserving and, in the case of Federal Learning, makes the best use of unlabeled data. Our method provides a scalable, privacy preserving, and practical solution to medical imaging that suffers from a lack of annotated data.

Keywords. EKDFSSL, Federated Learning (FL), Semi-Supervised Learning (SSL), Federated Semi-Supervised Learning (FSSL), Knowledge Distillation (KD), Pneumonia Detection, Chest X-ray, Privacy-Preserving AI, Consistency Regularization (CR), Ensemble Learning (EL)

1. INTRODUCTION

The rapid growth of artificial intelligence in healthcare has made medical image analysis a valuable asset for supporting timely and accurate diagnoses of various medical conditions. However, when considered in conjunction with the existing limitations of imaging modalities, the analysis of medical images continues is hampered by the limited availability of datasets and continual concerns for patient privacy.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Pneumonia diagnosis from chest X-ray images presents multiple challenges. As well as having to pay expert radiologists to review the images, the provision of labeled examples is extremely costly. Moreover, both the Health Insurance Portability and Accountability Act and the General Data Protection Regulation present constraints on sharing data between different medical institutions. Additionally, traditional deep learning systems face serious obstacles in privacy, communication costs, and practicality in operating in real world healthcare systems, due to a reliance on centralized data collection. Because of the issues outlined above, there is a compelling case for the development of learning systems that are privacy preserving and adaptable to both labeled and unlabeled data.

In the last few years, the ability to collaboratively develop machine learning models without the need for data sharing has resulted in federated learning becoming one of the most popular privacy preserving frameworks. However, despite the benefits of this novel approach, most federated learning techniques are designed with the assumption that all involved parties have access to some form of labeled data. However, this is most definitely not the case as most healthcare institutions will tend to have predominantly unlabeled data, like chest X-ray images, with only a small volume of data that is labeled. This means a considerable challenge for the machine learning community is to develop systems which are privacy preserving and address the above imbalance while providing high quality outcomes.

To tackle the above challenge, we propose a framework called EKDFSSL (Ensemble Knowledge Distillation for Federated Semi-Supervised Learning) that detects pneumonia from chest X-ray images. Our approach combines consistency regularization and ensemble knowledge distillation in a federated learning environment. Herein, the global model provides guidance to client models, which helps promote the internal models to predict consistently for different augmentations of the same unlabeled data. This stabilizes the learning process and protects against the global model's performance worsening. On the server side, a knowledge distillation ensemble is formed to combine the client model predictions and facilitate learning with limited labeled data while combating overfitting.

There are several aspects of this work that advance the state of the art. The first is the design of a framework for federated semi-supervised learning that operates in a realistic environment where all labeled data exists solely on the server, while the clients are equipped solely with unlabeled data. The second is augmentations-based consistency regularization technique in a global model to elevate the learning process of clients. The third is the incorporation of ensemble knowledge distillation to combine the models of clients to improve the overall model. The fourth is the enhancement of privacy due to minimal data transfer. Finally, we developed a web-based (Flask) interface to conduct secure and easy user interactive pneumonia screening.

This work introduces a novel pneumonia detection system that is both scalable and considers user privacy, and that has the potential to be used in real world healthcare settings.

2. LITERATURE SURVEY

2.1. Ensemble Knowledge Distillation for Federated Semi-Supervised Image Classification

This work presents a federated semi-supervised learning framework for image classification where the labels exist at the server. Here, the global model acts as a teacher in a consistency

regularization framework to enhance learning from unlabeled data at the client side. To counter the overfitting problem due to the scarce labeled data, an ensemble knowledge distillation loss is used for server-side retraining. The performance of the framework is evaluated on several federated datasets such as Fashion-MNIST, SVHN, CIFAR-10, and CIFAR-100 and the results outperformed existing frameworks such as FedMatch, FedRGD, FedMix, and GDST. The framework is designed to achieve high accuracy for both IID and non-IID data. The current work is also limited by the assumption that all clients possess unlabeled data whereas all the server has is labeled data, an assumption that may not hold in many real world cases [1].

2.2. Clients Help Clients: Alternating Collaboration for Semi-Supervised Federated Learning

One way to mitigate communication costs while dealing with data heterogeneity in semi-supervised federated learning involves alternating collaboration. Similarity-based clustering helps to form groups of clients that deal with comparable data distributions to generate robust pseudo-labels. To counteract bias from non-IID data, dissimilarity-based clustering is employed. To optimize the labeling and support balancing, a reinforcement learning strategy is incorporated. Compared to baseline methods, the proposed method demonstrated an accuracy increase, along with a reduction in communication costs. Although clustering at each training round does not account for the variability in client computational resources, it does introduce additional computation costs [2].

2.3. Federated Semi-Supervised Learning for Attack Detection in IIoT

A novel federated semi-supervised scheme for intrusion detection has been proposed for the Industrial Internet of Things (IIoT). Using federated averaging, local autoencoders are first trained on the available unlabeled data, and then the local models are aggregated at the server. The server then provides labeled data and trains the added classification layers. Detection accuracy is achieved while maintaining privacy through the scheme. The sparse labeled datasets and communication costs used in the experiments illustrate the high detection accuracy and performance of the proposed scheme. The lack of convolutional layers in the proposed design may prevent the learning and detection of complicated spatial intrusion patterns. In this design, the thresholds used for pseudo-labeling are fixed and may not be optimal throughout the training process [3].

2.4. Knowledge-Driven Federated Learning for AIS Signal Identification

The proposed approach integrates knowledge-based and semi-supervised Federated Learning (FL) methods to address classification of AIS signals under the constraint of unbalanced labeling. The approach uses an HSIC-based similarity consistency to align feature representation across clients and a voting-based method for the generation of pseudo-labels to enhance the quality of the labels. The approach is able to attain a high level of classification accuracy across disparate domains, and thus is able to handle imbalanced data. One of the drawbacks is that the computation of HSIC requires parameter tuning and the voting method increases the communication burden due to additional model exchanges [4].

2.5. Heterogeneity-Aware Semi-Supervised Federated Learning

The proposed approach deals with the different levels of computation and communication capability of clients by incorporating them into the clustering process to minimize the effect

of slow clients. Using a Multi-Armed Bandit approach, different levels of training iterations are implemented for the different clusters, which improves the overall process. The results showed a significant improvement in convergence and a reduction in the communication burden. One of the drawbacks of the approach is that the framework is based on the assumption that heads of clusters are trustworthy, which is not always the case, and that adaptive optimization increases the burden on computation at the server [5].

2.6. Privacy-Preserving Generative Network for Smart Grids

A framework for anomaly detection in smart grids has been developed that combines generative networks with semi-supervised federated learning. This framework is built on the combination of temporal convolutional blocks and the geometric median for aggregation under the non-IID assumption. The framework is capable of having a high degree of accuracy in classifying both binary and multi-class examples. One of the challenges with this approach is that the use of generative adversarial networks (GANs) is known to produce unstable solutions and can cause mode collapse. Moreover, more layers in the architecture increase the computational burden [6].

2.7. Robust Semi-Supervised Federated Learning for UAV Image Recognition

An image recognition framework, built with a robust semi-supervised federated learning approach, is designed for systems that utilize drones. This framework is flexible to the models of the labels, either at the client or at the server, and incorporates a parameter mixing strategy to integrate supervised, unsupervised, and global models. This framework also utilizes a frequency-based aggregation approach to remedy the non-IID assumption and employs a hyperparameter self-adjustment strategy. Despite the framework having stable performance across multiple conditions, the main drawback is the heavy computational burden and memory usage caused by the large number of models [7].

2.8. Random Sampling Consensus Federated Semi-Supervised Learning

To improve model reliability in non-IID situations, a federated semi-supervised learning method that uses random sampling consensus has been developed. By randomly selecting clients to generate sub-models and aggregating the sub-models using a distance-based weighting method, the method is able to focus on the reliable models and optimize the performance to a greater extent. This method has been experimentally found useful with a wide variety of datasets and numbers of clients. The random sampling approach may ignore valuable client inputs. Additionally, the distance-based weighting method has a high computational complexity that will increase with the number of clients [8].

2.9. Advanced Model-Contrastive Federated Learning via Knowledge Distillation

MOON-KD is the name of an advanced federated learning method that combines model-contrastive learning and knowledge distillation in the aim of attaining improved performance given non-IID data. The global model, in this context, functions as a teacher, instructing local models on the appropriate changes to be made via the distillation loss, which is integrated with the supervised contrastive loss, in a given objective function. The alignment of local models with the global model is, thus, retained, and relative model performance is improved. In particular, it has been found, in experimental work carried out using the CIFAR-10 dataset, that MOON-KD outperforms the baseline methods of MOON and FedAvg, and that the distillation and contrastive losses, in particular, define the

boundaries of the overall model performance. There is no personalization for the client within this method, and there may be a large communication burden as there is an exchange of large model parameters. There are proposed method improvements that will fall within the confines of model compression, privacy, and more dynamic aggregation methods [9].

2.10. Federated Learning Algorithm Based on Knowledge Distillation

FedDistill is an approach using Knowledge Distillation to tackle issues of disparate model architectures and data distributions at client sites in Federated Learning (FL). Here, clients store both a Personal and a Global model. The Personal model acts as a teaching model that distills knowledge to the Global model, enabling learning across diverse model architectures. Evaluation of the approach on several datasets, like MNIST and CIFAR-10, shows both local and global model accuracy gains. The approach also brings a communication cost advantage due to the many local updates it allows prior to model aggregation. However, this approach still suffers from challenges of the classical averaging frameworks and is sensitive in performance to the local training iteration ceiling. It also requires tuning to understand the ideal training parameters for local iterations [10]. The model cannot effectively be used for real time detection due to high computational demands. The model's complexity is due to additional layers, high parameter counts, and intricate MRS-YOLO feature extraction. This all results in a significant increase in memory utilization [10].

3. METHODOLOGY

This study aims to develop the framework for building a privacy-preserving pneumonia detection model from CXR images using the Federated Semi-Supervised Learning (FSSL) approach. The foundation of the proposed framework is the EKDFSSL, which combines client-side unsupervised learning and server-side supervised learning.

In the first step, the chest X-ray images are distributed across client nodes, which in this case are assumed to be hospitals and diagnostic centers. Each of these client nodes focuses on a large amount of unlabeled data, while the central server holds a small amount of labeled data. This situation is typical in practice, where there is a considerable amount of unlabeled data and a limited amount of sensitive labeled medical data.

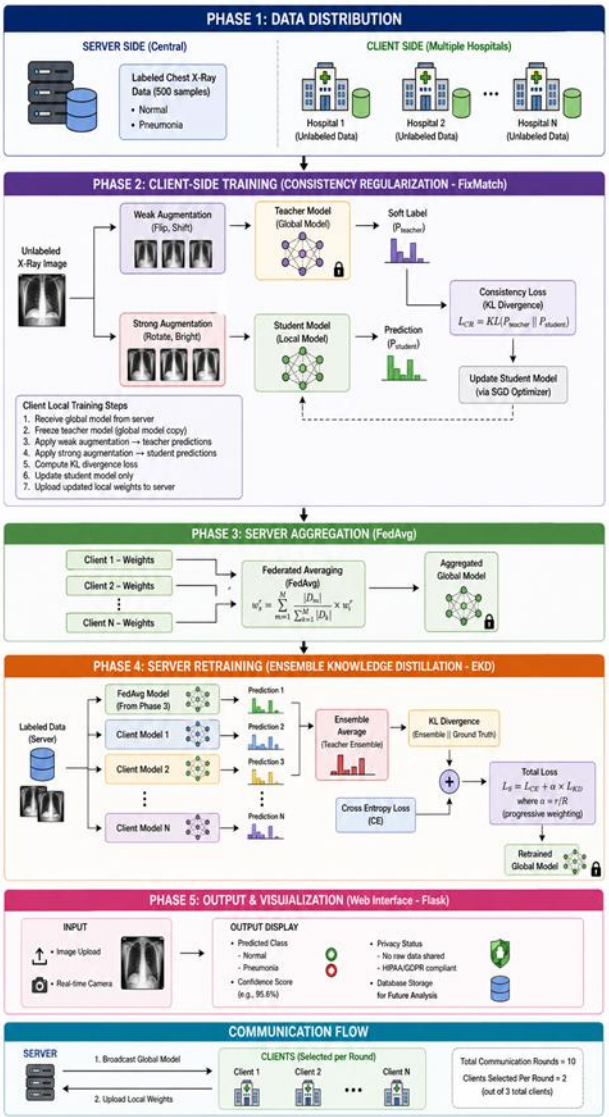
A global model is first created at the server and communicated to a subset of selected clients. Each client is then able to perform a training step on their unlabeled data. One approach to learning from unlabeled data is through the use of consistency regularization. In this framework, two different augmented versions of the same data sample are used. One is a weak version and the other is a strong version, and the model is trained to produce consistent output for both. The global model acts as a teacher to the client model, which learns to develop stable, meaningful representations.

Once local training is complete, each client communicates the updated model parameters to the server. The server then uses the Federated Averaging (FedAvg) algorithm to develop an enhanced global model. Knowledge is combined in the aggregation process, which supports privacy, as it does not involve the communication of any raw data.

The server then conducts a supervised re-training phase on its collected labeled dataset. Here, the model is optimized using a combination of the cross-entropy loss and the ensemble

knowledge distillation loss. In the distillation process, predictions from several client models serve as an ensemble teacher and offer a more informative supervision to the global model.

The international model is then assessed using a test data set. The complete system is implemented via a web interface built with Flask. This enables health care workers to use the model for pneumonia screenings with the assurance of data safety and privacy.



3.1. Data Distribution

For the system we propose, gross examination of chest X rays data is shared among a central server and several clients (hospitals). The server stores a small labeled data set of normal and pneumonia cases, while each client stores a large data set of unlabeled chest X ray images. This distributed system imitates most real world medical situations where labeled

data is of a small volume, sensitive, and contains private medical information. It contains the unrefined patient data and creates a variety of situations within the environment. The system is designed to better the performance of the system in the real world.

3.2. Client-Side Training using Consistency Regularization

For the client based training, we use the semi-supervised learning based on the consistency regularization (FixMatch) on the unlabeled data. For every input image, we generate two images, one with weak augmentation and one with strong augmentation. The weakly augmented image is sent to the global model, which in this case acts as a teacher to generate soft pseudo-labels. The local model (the student) is then tasked with the strongly augmented image and generates a prediction.

The consistency loss along with a Kullback-Leibler (KL) divergence helps bring the teacher and student outputs closer. The student model is then fine-tuned with an optimizer, such as stochastic gradient descent (SGD). This approach helps the model learn from unlabeled data and keeps the predictions consistent for the different transformations applied to the data.

3.3. Server Aggregation using Federated Averaging (FedAvg)

Once the local training is complete, the updated model weights from a client are communicated to the central server. The server then applies the FedAvg algorithm to aggregate the models. In this method, the contribution made by a client is dependent on the client's dataset size. The resulting global model is enhanced with the knowledge from all the clients, and the privacy of the clients is maintained as the raw data from the clients is not accessed.

3.4. Server-Side Retraining using Ensemble Knowledge Distillation (EKD)

The server focuses on model performance enhancement by retraining with its labeled data. This stage utilizes Ensemble Knowledge Distillation (EKD). This method generates predictions for the same input from the global and several client models, which are then combined through an ensemble averaging method to create a more powerful supervisor signal.

Training is performed using a combination of cross-entropy loss (for labeled data) and knowledge distillation loss (based on ensemble predictions). During this training, control for each loss is introduced. This method significantly improves model performance when labeled data is insufficient.

3.5. Output and Visualization

The server focuses on model performance enhancement by retraining with its labeled data. This stage utilizes Ensemble Knowledge Distillation (EKD). This method generates predictions for the same input from the global and several client models, which are then combined through an ensemble averaging method to create a more powerful supervisor signal.

Training is performed using a combination of cross-entropy loss (for labeled data) and knowledge distillation loss (based on ensemble predictions). During this training, control for each loss is introduced. This method significantly improves model performance when labeled data is insufficient.

4. ARCHITECTURE

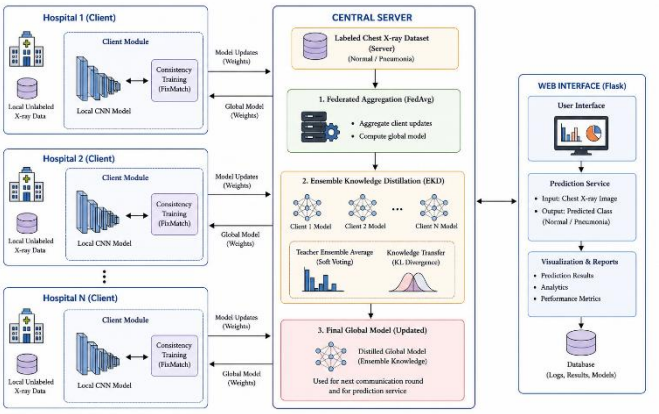


Fig 1: Architecture Diagram

The server enhances model performance by retraining with the labeled dataset. The process uses Ensemble Knowledge Distillation (EKD). Given the same input, the global model and various client models generate predictions, which, after being synthesized with an ensemble averaging technique, provide a more robust supervisory signal.

The training process uses a combination of cross-entropy loss (for labeled data) and knowledge distillation loss (derived from ensemble predictions). A progressive weighting factor assigns balance to the losses during the training process. The model's generalization and accuracy are improved with the limited labeled data.

4.1. Client Data Module

The system targets a distributed architecture with various client nodes (hospitals) as data sources. Each client has a collection of locally stored chest X-ray images, most being unlabeled, for data privacy. Furthermore, the central server stores a small labeled dataset with normal and pneumonia cases.

We intentionally split the data to mimic the challenge in the medical field, where there is limited labeled and sensitive data, but there exists a considerable volume of unlabeled data across various medical institutions. The module ensures that data remain in their respective clinics and are only utilized for local computation.

4.2. Client-Side Training Module

This module manages the local model training for each client. Relevant features from chest X-ray images are extracted using a convolutional neural network (CNN). To enhance model performance with available unlabeled data, the system employs a consistency-based learning approach due to the limitation of labeled data at the client side.

The global model is sent from the server to the clients, while the clients make revisions to the global model using their local datasets. The training process is designed to make the model produce robust and consistent predictions, even when the input data vary

significantly. When the training process is complete, the model is packaged, and only the revised model parameters are sent to the server.

4.2.1. Feature Extraction Module

The central server coordinates the learning process. When the clients update their models, they send the model updates to the central server. The server updates the model without needing to see the raw data. The server uses a technique called Federated Averaging to combine model parameters. This technique allows the server to learn from the clients and keeps the data private. After the model is updated, the server sends the global model to the clients to continue the training process. This framework supports a learning process that is both iterative and collaborative.

4.2.2. Server-Side EKD Module

To improve the global model, the server adds an EKD Module. In this case, the EKD Module combines the predictions of multiple client models and the aggregated global model.

The model is retrained with the labeled data the server has to improve the model's accuracy and its ability to generalize. The system is able to optimally combine both labeled and unlabeled datasets. The system also shows an increased robustness to unstandardized medical image data.

4.2.3. Output and Evaluation Module

The final global model is integrated into a web-based system where users can submit chest X-ray images. The system processes the images and classifies the cases as normal or pneumonia with confidence score. The system is evaluated with traditional classification metrics of accuracy, precision, recall and F1-score. These metrics are benchmarks of the system's effectiveness and reliability in real world clinical applications. Also, the performance is evaluated based on the number of iterations of the training rounds to check if the model is improving.

5. DATASET DESCRIPTION

The dataset is composed of chest X-ray images (normal cases and pneumonia cases) collected from open sourced medical imaging. The dataset is purposefully arranged to comply with the federated learning framework. There is a small amount of labeled data at the central server and large amounts of unlabelled data at the various client locations (i.e. hospitals). This data arrangement imitates real world clinical situations where there is small amount of labeled data and large amount of unlabelled data readily available.

5.1. Data Sources

The chest X-ray data was obtained from sites like Kaggle that host open source medical data. These data repositories contain X-ray data that are specifically collected for pneumonia detection.

The server's dataset contains labeled data that are clearly classified to facilitate supervised learning and model assessment. These data repositories are highly regarded in academic research and therefore can be used to evaluate and benchmark model's performance.

5.2. Dataset Composition

There are two primary classes in this dataset.

Normal class contains Chest X-ray images which are devoid of infectious signs.

The Pneumonia class contains images that depict signs of lung infection.

In the described system, the labeled dataset is centralized, and the client devices hold unlabeled images. This allows for combining the supervised learning at the server and semi-supervised learning at the client, respectively.

5.2.1. Data Annotation

The labeled dataset contains annotations that are at the image level and classify the image to either the Normal or Pneumonia class. These annotations are utilized in the training and evaluation at the server.

5.2.2. Data Preprocessing

Before the images are utilized for training, they undergo some basic preprocessing to ensure uniformity. The images are resized, the pixel values are normalized and converted to the appropriate image input for the model.

These help to improve the consistency of the training and ensure the model is able to learn the important and the correct patterns from the data.

5.2.3. Data Augmentation

For the purpose of creating a robust model, various data augmentation techniques are employed in the training, which may include flipping, rotation, scaling, and brightness.

These changes allow the model to become more flexible and adaptive to the real-world challenges posed by medical imaging, such as variable image quality, fluctuating orientation, and poor or imperfect lighting.

5.2.4. Dataset Split

For evaluation purposes, the labeled dataset is partitioned into training, validation, and testing sets to promote efficient learning and allow for a more accurate assessment of performance. The model training employs roughly 80% of the overall dataset, where the remaining 20% is allocated for validation and testing. The training dataset allows the model to learn critical distinguishing attributes of the chest X-ray images. The validation dataset is used to control the learning process to prevent overfitting.

After training, the model is tested with the testing dataset to assess the performance of the model beyond the constraints of the training dataset. This is the most accurate indication of themodel's ability to generalize in the context of a real-world medical diagnosis. The dataset splitting strategy promotes a fair assessment and enhances the reliability of the results from the experiments.

6. RESULTS

We assess the federated semi-supervised medical image classification framework for chest X-ray images dispersed across multiple hospital clients. Each hospital develops its own local CNN model on the local chest X-ray data. To avoid sharing local chest X-ray data, hospitals send the learned model weights to the central server. This maintains data privacy and security during the training process.

The central server is essential to enhance the overall model. After receiving model updates from all the clients, the central server uses the Federated Averaging (FedAvg) algorithm to construct a global model. Furthermore, Ensemble Knowledge Distillation (EKD) is applied to improve the classification performance by bridging the knowledge gap between client models, thereby achieving efficient distributed learning.

We demonstrate that our DenseNet121-based model has a superior performance on medical image classification when compared to the ResNet model. DenseNet121 uses a principal design that allows the model to learn which features are necessary to classify images, thus providing a better mechanism to identify and learn the important features of chest X-Ray images. Our model has an accuracy of almost 83%, which demonstrates that our model is reliable in distinguishing between normal and abnormal chest X-Ray images.

The training process demonstrates stability with multiple rounds of communication between the client and server. Training loss is decreasing over time, and so the model is learning. The validation loss, similarly, shows stability, indicating that the model is able to generalize to unseen data, with no significant overfitting.

The usability of the system is further enhanced with the addition of the Flask-based web interface, which allows users to upload chest X-ray images and receive predictions. The system also maintains a database of prediction logs, model info, and prediction-related reports. The system provides an efficient, safe, and privacy-preserving way to implement federated medical image classification with deep learning.

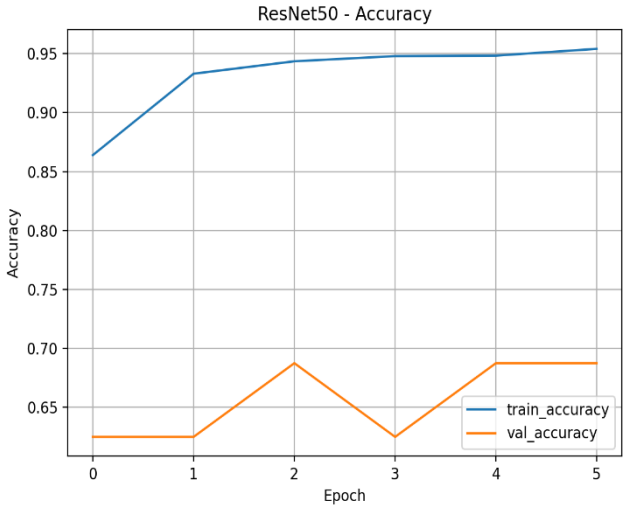


Fig 2: ResNet Accuracy

The performance of the ResNet model during training was inferior and it was not able to accomplish chest X-ray classification results on par with DenseNet121. There was some improvement to training accuracy, but it was slow and inconsistent. The validation accuracy was also low, but with more frequent fluctuations, which shows poor generalization to chest X-ray images.

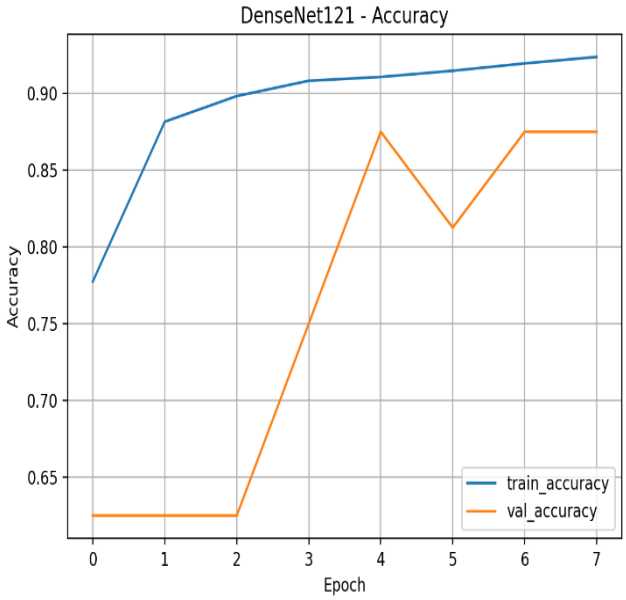


Fig 3: DenseNet Accuracy

DenseNet121 reached an impressive nearly 87% validation accuracy, while ResNet struggled to surpass 70% accuracy for most of the training rounds. It was fairly evident in the training and validation results that ResNet was not as suitable for the medical image classification task, while DenseNet121 offered better learning, feature extraction, and consistency for effective pneumonia detection.

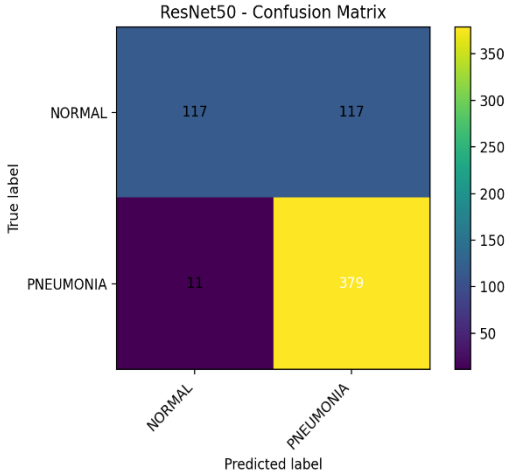


Fig 4: Cofusion Matrix of ResNet

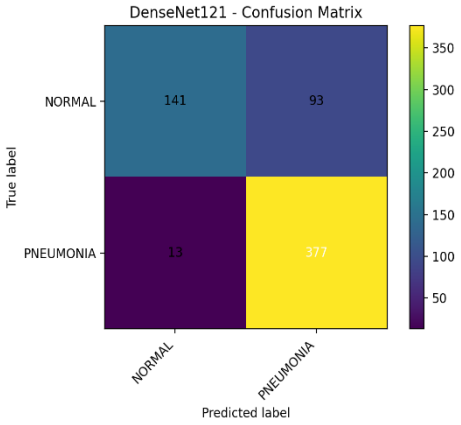


Fig:5 Cofusion Matrix of DenseNet.

The confusion matrix illustrates that DenseNet121 was superior to ResNet in classifying pneumonia in chest X-ray images. DenseNet121 classified the majority of pneumonia instances and had a low number of false-negative instances, making it an excellent choice for medical detection tasks. In comparison, ResNet had an increased number of false-positive instances and was unable to distinguish pneumonia images from normal images. The overall image classification results depict that DenseNet121 is the most accurate and reliable of the two for the task of classifying pneumonia in chest X-rays.

7. CONCLUSION

This paper proposes a system for automated pneumonia detection in chest X-ray images using deep learning. The system utilizes DenseNet121 for feature extraction and classification as a result of the benefits that the architecture and transfer learning provide. The system also implements knowledge distillation to simplify the model while retaining the diagnostic strength. The results of the experimental evaluations suggest that the model is capable of maintaining a validation accuracy of 87% and a pneumonia detect rate of 96.7%, indicating that the model is capable of accurately identifying pneumonia cases with a limited number of missed cases.

Training and validation results show good convergence during training, with decreasing loss values and successful feature learning. Training and validation accuracy values show good generalization, with limited overfitting. The analysis of the confusion matrix shows the model is efficient, with only a small number of pneumonia cases being misclassified. While normal case classification accuracy is moderate, the model's priority is on the classification of pneumonia cases, which is important in medical diagnosis, to reduce the chances of missing a pneumonia infection.

The method proposed shows a high degree of applicability in the medical field in that it aids the radiologist in the early detection of pneumonia and in the lessening of the overall workload associated with diagnosis. The framework will be useful in future research to be extended to the classification of other respiratory diseases, COVID-19 and tuberculosis, and the presence of lung opacity in X-ray images. In addition, lightweight distilled models may enable portable edge devices to be used for diagnosis in real time, and combine with explainable AI methods to improve clinical readability by visually annotating areas of pneumonia on chest X-ray images.

8. FUTURE WORK

The proposed system shows strong performance in detecting pneumonia from chest X-ray images. There are, however, additional features that can be added to the system to increase its capability and applicability in a clinical setting. The system currently classifies images as Normal and Pneumonia. However, the system can be developed to classify additional respiratory illnesses such as COVID-19, tuberculosis, lung opacity, pneumothorax, and pulmonary fibrosis. This would develop a more comprehensive system for the classification of medical images in the screening of respiratory illnesses.

The classification system can be further improved through the development of advanced teacher-student frameworks in knowledge distillation. More advanced heterogeneous distillation frameworks, where the teacher and student are of different architectures, can help decrease the computational burden of the classification system, while still maintaining classification performance. The quality of knowledge distillation can also be improved by the use of adaptive temperature scaling and attention transfers, among other techniques.

The proposed system is also adaptable for low-resource healthcare settings, such as rural hospitals, emergency care systems, and mobile healthcare units. Lightweight distilled student models can be deployed to the edge for real-time diagnosis with a lower computational burden. Future work can also incorporate dynamic chest imaging and

fluoroscopy with a focus on improving temporal features with more consistency in the diagnosis.

9. REFERENCE

[1] E. Shang, H. Liu, J. Zhang, R. Zhao, and J. Du, "Ensemble knowledge distillation for federated semi-supervised image classification," Tsinghua Science and Technology, vol. 30, no. 1, pp. 112-123, Feb. 2025.

[2] Z. Jiang, Y. Xu, H. Xu, Z. Wang, and C. Qiao, "Clients help clients: Alternating collaboration for semi-supervised federated learning," in Proc. IEEE 40th Int. Conf. Data Engineering (ICDE), 2024, pp. 1847-1860.

[3] O. Aouedi, K. Piamrat, G. Muller, and K. Singh, "Federated semisupervised learning for attack detection in industrial internet of things," IEEE Trans. Ind. Informat., vol. 19, no. 1, pp. 995-1005, Jan. 2023.

[4] Q. Tian, S. Wang, and M. Wang, "FSLBSC: Knowledge-driven federated learning for unbalanced labeling AIS signal identification," IEEE Trans. Intell. Transp. Syst., 2025.

[5] Y. Xu, Z. Yao, H. Xu, Y. Liao, Y. Zhu, and X. Fu, "Heterogeneity-aware semi-supervised federated learning with client-to-client collaboration," IEEE Trans. Mobile Comput., 2025.

[6] M. Abdel-Basset, N. Moustafa, and H. Hawash, "Privacy-preserved generative network for trustworthy anomaly detection in smart grids: A federated semisupervised approach," IEEE Trans. Ind. Informat., vol. 19, no. 1, pp. 995-1005, Jan. 2023.

[7] Z. Zhang, S. Ma, Z. Yang, Z. Xiong, J. Kang, Y. Wu, K. Zhang, and D. Niyato, "Robust semisupervised federated learning for images automatic recognition in internet of drones," IEEE Internet Things J., vol. 10, no. 7, pp. 5733-5746, Apr. 2023.

[8] X. Liang, Y. Lin, H. Fu, L. Zhu, and X. Li, "RSCFed: Random sampling consensus federated semi-supervised learning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 1-10.

[9] Y. Chen, "Advanced model-contrastive federated learning via knowledge distillation," in Proc. IEEE 3rd Int. Conf. Control, Electron. Comput. Technol. (ICCECT), 2025.

[10] D. Jiang, C. Shan, and Z. Zhang, "Federated learning algorithm based on knowledge distillation," in Proc. Int. Conf. Artif. Intell. Comput. Eng. (ICAICE), 2020.

[11] S. Caldas, J. Konecny, H. B. McMahan, and A. Talwalkar, "Expanding the reach of federated learning by reducing client resource requirements," arXiv preprint arXiv:1812.07210, 2018.

[12] C. Xie, S. Koyejo, and I. Gupta, "Asynchronous federated optimization," arXiv preprint arXiv:1903.03934, 2019.

[13] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in Proc. IEEE Int. Conf. Communications (ICC), Shanghai, China, 2019, pp. 1-7.

[14] D. Yang, Z. Xu, W. Li, A. Myronenko, H. R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang, et al., "Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan," Med. Image Anal., vol. 70, p. 101992, 2021..

[15] A. Y. Zhu, Y. Liu, J. J. Q. Yu, and X. Yuan, "Semi-supervised federated learning for travel mode identification from GPS trajectories," IEEE Trans. Intell. Transport. Syst., vol. 23, no. 3, pp. 2380-2391, 2022.

[16] Y. Zhao, H. Liu, H. Li, P. Barnaghi, and H. Haddadi, "Semi-supervised federated learning for activity recognition," arXiv preprint arXiv:2011.00851, 2020.

[17] W. Jeong, J. Yoon, E. Yang, and S. J. Hwang, "Federated semi-supervised learning with inter-client consistency & disjoint learning," arXiv preprint arXiv:2006.12097v3, 2020.

[18] Z. Zhang, Y. Yang, Z. Yao, Y. Yan, J. E. Gonzalez, K. Ramchandran, and M. W. Mahoney, "Improving semi-supervised federated learning by reducing the gradient diversity of models," in Proc. IEEE Int. Conf. Big Data (Big Data), Orlando, FL, USA, 2021, pp. 1214-1225.

[19] Z. Zhang, S. Ma, J. Nie, Y. Wu, Q. Yan, X. Xu, and D. Niyato, "Semi-supervised federated learning with non-IID data: Algorithm and system design," in *Proc. IEEE 23rd Int. Conf. High Performance Computing & Communications; 7th Int. Conf. Data Science & Systems; 19th Int. Conf. Smart City; 7th Int.

[20] Z. Zhang, S. Ma, Z. Yang, Z. Xiong, J. Kang, Y. Wu, K. Zhang, and D. Niyato, "Robust semi-supervised federated learning for images automatic recognition in Internet of drones," IEEE Internet Things J., vol. 10, no. 7, pp. 5733-5746, 2023.

[21] X. Liu, L. Zhu, S. T. Xia, Y. Jiang, and X. Yang, "GDST: Global distillation self-training for semi-supervised federated learning," in Proc. IEEE Global Communications Conf. (GLOBECOM), Madrid, Spain, 2021, pp. 1-6.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Automated Glacier Boundary Delineation and Change Detection Using U-Net and CNN on Sentinel-1 SAR Imagery

Jagdish Kapadnis[1], Dr. Avinash Dhole[2], Jayesh J. Pandey[3], Kanishk T. Vadge[4], Prerana R. Thok[5]

[1]*Asst. Prof., Computer Engineering, PVGs COE, Nashik, jagdishkapadnis@gmail.com*

[2]*Asso. Prof., Computer Engineering, SSSTc, Bhilai, dhole1072@gmail.com*

[3]*Computer Engineering, PVGCOE-SSDIOM, Nashik, jayeshpandey754@gmail.com*

[4]*Computer Engineering, PVGCOE-SSDIOM, Nashik, yash.vadge04@gmail.com*

[5]*Computer Engineering, PVGCOE-SSDIOM, Nashik, prernathok7911@gmail.com*

Abstract.

In worldwide effect of climate change roughly indicated by Glaciers. So it is most important that, accurate mapping of their boundaries are essential for to measure environment impact. The proposed research work on Deep Learning (DL) based framework for automated glacier detection using optical satellite images and corresponding ground-truth. The work is divided into two semantic segmentation models, first model depends upon classical U-Net architecture with an Encoder/Decoder structure with skip connections and second model a custom Convolutional Neural Network (CNN) designed to feature extraction was developed and evaluated. All the models are trained using Glacier-Lake PNG data-set. Overall Experimental results clearly indicates that, the U-Net model outperforms the CNN, achieving 97.74% test accuracy, along with higher overlap metrics, Dice coefficient and Intersection over Union (IoU). Final visual analysis of predicted segmentation masks further demonstrates that, improved boundary delineation capability of U-Net, in complex regions containing debris and shadows. Finally, the proposed framework demonstrates an efficient and scalable approach for glacier monitoring and it will support large-scale climate change studies.

Keywords. Glacier Mapping, Sentinel-1 SAR, Semantic segmentation

1. INTRODUCTION

Glaciers are the prime indicators of climate change, some are the common indicators are sea-level rise, water resource variability, and hazards like glacial lake outburst floods. Traditional method of manual mapping glaciers from satellite imagery is very time consuming [1], subjective, and difficult in complex terrains. Limitation of remote sensing offers large-scale monitoring, but distinguishing glacier ice from debris and snow remains challenging. The proposed research offers Deep Learning (DL) model like semantic segmentation models like CNNs and U-Net, has improved automated glacier detection. Mainly here it compares U-Net with a baseline CNN [2] using optical satellite images and binary masks, aiming to enhance segmentation accuracy and enable multi-temporal glacier change analysis.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. BACKGROUND AND RELATED WORK

Nowadays, glacier monitoring is increasingly important due to climate change and its impacts on global sea levels, water resources, and surrounding ecosystems. Traditional satellite remote sensing [3], particularly Landsat and Sentinel imagery, provides valuable high-resolution observations for monitoring glacier dynamics over large areas. However, accurately delineating glacier boundaries remains challenging because of complex mountainous terrain and varying environmental conditions [4]. Glacier mapping using the Normalized Difference Snow Index (NDSI) combined with digitization techniques [5] can provide reliable results under favorable conditions but performs poorly in areas with debris-covered ice, shadows, and seasonal snow. Moreover, manual mapping over large regions is labor-intensive and time-consuming. To address these limitations, deep learning (DL) techniques have demonstrated strong potential for remote sensing applications, particularly semantic segmentation [6]. Convolutional neural networks (CNNs) and architectures such as U-Net can learn complex spatial patterns directly from imagery, enabling automated and accurate glacier boundary extraction. These capabilities offer improved efficiency and reliability for glacier monitoring systems [7].

3. METHODOLOGY

The proposed framework employs a deep semantic segmentation pipeline for automatic glacier detection from optical satellite imagery using an improved U-Net with a customized CNN. U-Net adopts an encoder-decoder architecture [8] to learn hierarchical features while preserving spatial information through skip connections. Multi-Head Self-Attention (MHSA) is incorporated at the bottleneck to capture long-range dependencies [9] between glacier regions and visually similar backgrounds, including snow, rocks, and clouds.

3.1. U-Net Architecture

The segmentation pipeline, illustrated in Figure 2, uses a symmetric encoder-decoder structure with skip connections [10]. The encoder processes 256×256×3 RGB images using 3×3 convolutions, ReLU, batch normalization, dropout, and max-pooling. The decoder uses transposed convolutions and skip connections to restore spatial resolution, followed by 1×1 convolution and SoftMax activation for binary glacier-background segmentation..

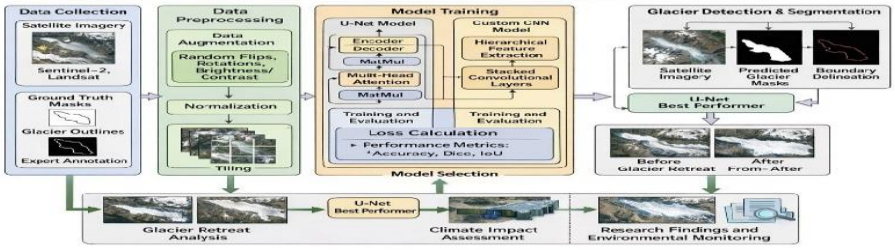


Figure 3.1.1 Glacier Segmentation Architecture

3.2. Custom CNN Architecture

The baseline CNN utilize the sequential convolutional layer for feature extraction, along with down sampling and up sampling for segmentation output. The details customized CNN U-Net architecture shown in figure 3.

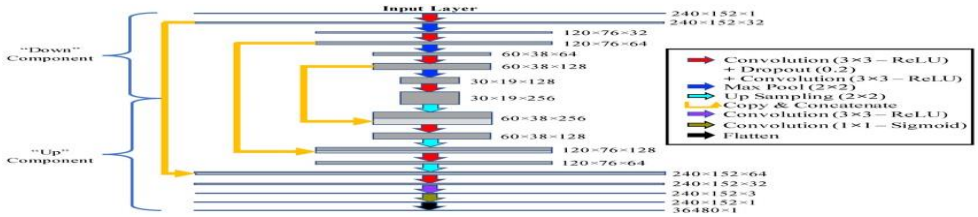


Figure 3.2.1. U-Net Architecture with down sampling, up sampling, skip connections, and convolutions for precise image segmentation.

3.3. System evaluation parameters

Input optical satellite image $\mathbf{I} \in \mathbb{R}H \times W \times 3$ (height H, width W, 3 RGB channels) predict a binary segmentation mask $\mathbf{M} \in \{0,1\}H \times W$. The deep learning model $f\theta$ approximates a mapping: $I \in \mathbb{R}^{H \times W \times 3}$. The segmentation model f_θ(U-Net or CNN) learns a mapping from the input image to a pixel-wise probability map for two classes: glacier and background. The predicted binary segmentation mask is obtained by assigning each pixel to the class having the highest probability:

$$\hat{M} = \arg\max_c \hat{Y}_{(:,:,c)} \tag{1}$$

The training is done through the categorical cross-entropy loss over N training samples:

$$L(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{h,w,c} Y_i^{(h,w,c)} \log\left(\hat{Y}_i^{(h,w,c)}\right) \tag{2}$$

Where $Y_i$denotes the one-hot encoded ground-truth mask corresponding to glacier and background classes. To improve generalization, increase the variety of training samples, several augmentation techniques were applied during training. Brightness augmentation is applied as

$$I' = I + \delta_b, \delta_b \sim U(-0.1, 0.1) \tag{3}$$

while contrast adjustment is defined as

$$I' = (I - 0.5)\alpha + 0.5 \tag{4}$$

where $\alpha \sim U(0.8, 1.2)$. Pixel values are clipped to ensure that $I' \in [0,1]$ For multi-temporal glacier analysis, segmentation masks are generated independently for satellite images acquired at times $t_1$and t_2:

$$\hat{M}_{t_1} = f_\theta(I_{t_1}), \hat{M}_{t_2} = f_\theta(I_{t_2}) \tag{5}$$

A change map is computed using binary differencing:

$$\Delta M = \hat{M}_{t_2} \oplus \hat{M}_{t_1} \tag{6}$$

where $\oplus$denotes the XOR operation that highlights pixels whose class label has changed. The glacier area change is denoted as ΔA where A_p indicates the ground area represented by every pixel. The proposed approach enables quantitative assessment of glacier retreat.

$$\Delta A = A_p\left(\sum \hat{M}_{t_1} - \sum \hat{M}_{t_2}\right) \tag{10}$$

4. EXPERIMENT RESULTS

Both models were trained for 50 epochs using batch size 8, Adam optimizer, and categorical cross-entropy (Table 1). Data were divided into training (70%), validation (20%), and testing (10%). U-Net outperformed Custom CNN, achieving 97.74% accuracy, 96.8% Dice, and 93.8% IoU, indicating superior glacier-region overlap and boundary accuracy.

Table 1: Comparative chart of U-Net and Custom CNN on Test Set

Model	Accuracy (%)	Dice Coefficient	IoU	Precision (Glacier)	Recall (Glacier)
Custom CNN	93.24	0.912	0.838	0.895	0.931
U-Net	97.74	0.968	0.938	0.976	0.961

5. CONCLUSIONS

Finally, in proposed work, two segmentation models were tested and indicate that, U-Net outperformed the CNN, by achieving highest accuracy, Dice score and IoU over the custom CNN. This result is achieved by utilizing its encoder decoder design and connections that preserve spatial details. The pipeline contains pre-processing, data augmentation, training and enabling reliable tracking of the glaciers. Overall this scheme offers reliable solution over the traditional system to identify glacier retreats.

6. REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv preprint arXiv:1505.04597, 2015.

[2] F. Wang, S. Liu, L. Wang, C. Xu, and H. Meng, “Mountain glacier preservation with artificial interventions: A review,” *iScience*, vol. 29, no. 2, Art. no. 114744, Feb. 20, 2026, doi: https://doi.org/10.1016/j.isci.2026.114744

[3] Gupta C, Cryosphere and Ayurveda: A Holistic Framework for Glacier Preservation and Medicinal Plant Sustainability during the International Year of Glaciers (2025). J Ayu Int Med Sci. 2026;11(1):283-295.

[4] M. Huss, “On the feasibility of glacier preservation,” *Nature Water*, vol. 2, pp. 606–607, Jul. 2024.

[5] A. Diaconu et al., "Multi-Sensor Deep Learning for Glacier Mapping," arXiv preprint arXiv:2409.12034, 2024.

[6] C. Niemann et al., "Globally scalable glacier mapping by deep learning matches expert delineation accuracy," Nature Communications, vol. 15, 2024.

[7] X. Wang et al., "Mapping mountain glaciers using an improved U-Net model with CSE," International Journal of Digital Earth, vol. 15, no. 1, pp. 1–20, 2022.

[8] S. Nijhawan et al., "Glacier Boundary Mapping Using Deep Learning Classification over Bara Shigri Glacier in Western Himalayas," Sustainability, vol. 14, no. 20, 2022.

[9] R. Dietrich-Sussner *et al*., "Synthetic Glacier SAR Image Generation from Arbitrary Masks Using Pix2Pix Algorithm," *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Brussels, Belgium, 2021, pp. 4548-4551, doi: 10.1109/IGARSS47720.2021.9553853.

[10] D. J. Brinkerhoff, "Compatible Finite Elements for Glacier Modeling," in *Computing in Science & Engineering*, vol. 25, no. 3, pp. 18-28, May-June 2023, doi: 10.1109/MCSE.2023.3305864.

Biographies



Jagdish Kapadnis is an academician and researcher in the field of Computer Engineering**,** His research work has received 32 citations, an h-index of 3, and an i10-index of 1**.**



Dr. Avinash Dhole is an academician and researcher in the field of Computer Engineering**,** He has published several research papers in international journals and conferences.



Jayesh Pandey received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.



Kanishq Vadge received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.



Prerna Thok received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.

A Real-Time Smart Surveillance Framework Using Edge–IoT for Suspicious Activity Detection and Facial Emotion Analysis

Nidhi P. Nandankar, Shraddha R. Khonde

Department of Computer Engineering, Modern Education Society's Wadia College of Engineering, Pune- 411001, nidhinandankar25@gmail.com, shraddha.khonde@mescoepune.org

Abstract.

The growing need for intelligent, proactive surveillance has exposed the limitations of traditional CCTV systems, which rely heavily on continuous human monitoring and post-incident analysis. To address these challenges, this work presents an Edge-IoT-based real-time smart surveillance framework that integrates human activity recognition with facial emotion analysis using deep learning. The proposed system performs on-device processing on a Raspberry Pi, enabling local video analysis with low latency while preserving user privacy.Facial regions are detected using a lightweight Haar Cascade approach, followed by emotion classification with a CNN trained on the FER-2013 dataset. In parallel, human activities such as walking, running, and aggressive movements are analyzed using motion-based features. A decision-level fusion strategy integrates activity patterns with emotional cues such as fear and anger to accurately distinguish normal from suspicious behavior, thereby minimizing false alarms. Experimental results demonstrate reliable real-time performance on edge hardware, confirming the system's suitability for smart surveillance applications in public and institutional environments.

Keywords: Smart Surveillance; Facial Emotion Recognition; Suspicious Activity Detection; Convolutional Neural Network; IoT

1. INTRODUCTION

Traditional CCTV requires continuous human monitoring and has limited real-time intelligence [1]. Deep learning improves activity detection but often requires high computation or cloud processing, causing latency and privacy concerns [2–4]. Motion-based methods may produce false alarms, while emotion recognition is rarely integrated with activity detection [5–7]. This paper proposes a Raspberry Pi-based edge-IoT system combining activity and emotion recognition for accurate, private, and low-latency surveillance.

2. LITERATURE REVIEW

Existing methods provide good detection accuracy but face challenges such as high computational cost, false alarms, limited contextual awareness, and cloud dependency [10–18]. The key research gap is the lack of an integrated, real-time, privacy-preserving edge system combining activity and emotion recognition.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

The suggested framework adheres to an Edge-IoT-enabled intelligent surveillance paradigm capable of detecting human behaviors through joint analysis of both human actions and human facial emotions. The suggested approach differs significantly from traditional CCTV camera systems, which depend heavily on nonstop video recording and subsequent human supervision of the recorded videos

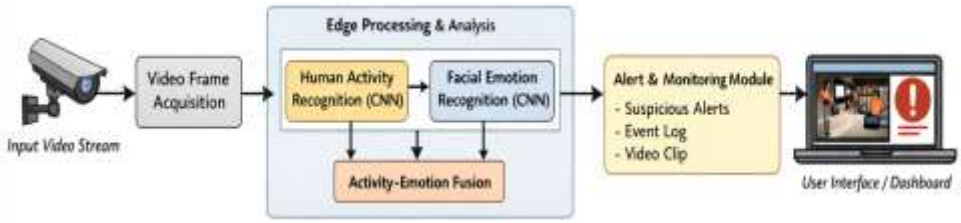


Figure 3.1. Architecture of the Proposed Edge-IoT system

Overall Framework

As shown in Fig.3.1 the architecture is organized into four functional layers:

- Video Acquisition Layer,
- Edge Processing and Feature Extraction Layer,
- Activity–Emotion Fusion and Decision Layer, and
- Alert and Monitoring Layer.

A Raspberry Pi is used as the main edge device, connected to the surveillance camera that will record video streams and conduct inference in real time.

Methodology

The concept of intelligent surveillance systems by leveraging Raspberry Pi with edge computing is introduced, employing the decision-level fusion mechanism of activity and facial emotion recognition using CNN-based architecture. This will enable real-time and efficient analysis without relying on cloud connectivity. The overall workflow of the proposed system consists of the following stages:

1. Video frame acquisition and preprocessing on Raspberry Pi
2. CNN-based human activity feature extraction
3. CNN-based facial emotion recognition
4. Decision-level activity–emotion fusion
5. Final suspicious activity decision and alert generation

3.2.1 Video Preprocessing and Frame Sequencing

Let the continuous video stream captured by the Raspberry Pi camera module or USB camera be represented as: V={F1,F2,…,Fn}(1) where F_i denotes the ith video frame. All captured video frames are resized and normalized to a size compatible with "lightweight" CNAs; that is, each individual frame will be processed in real-time by a Raspberry-Pi operating system when resized to a single height and width measurement (e.g., 1280x720).The final sliding window is then constructed by "sliding" that window over those k consecutive video frames:

$$S_t=\{F_t,F_{(t+1)},\ldots,F_{(t+k)}\} \quad (2)$$

Using sliding windows helps the model detect motion patterns over shorter periods of time, which allows for greater accuracy when detecting abnormal motion patterns based on prior observations alone. Such simplified temporal modeling is computationally efficient and well-suited for edge devices, while still offering meaningful activity context compared to more complex temporal learning frameworks

3.2.2 CNN-Based Human Activity Recognition

The human activity recognition unit employs an efficient CNN architecture for extracting motion information of spatial nature from the frame sequence S_t by means of the convolutional and pooling layers.

$$X_l = f(W_l * X_{l-1} + b_l). \quad (3)$$

(3) where W_l and b_l stand for the convolution and bias parameters of the l-th layer, * is a convolution operator, and f(·) is an activation function. The extracted features are used by the softmax classifier to assign an activity label from the list including, e.g., walking, running, fighting, and abrupt motion:

$$P(A_i \mid S_t) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (4)$$

where C refers to the total number of activity classes. Activities such as running or fast walking are considered ambiguous rather than suspicious in order to avoid false positives

3.2.3 CNN-Based Facial Emotion Recognition

Facial emotion recognition operates in parallel with activity detection. Face regions are first detected from video frames using a Haar Cascade or lightweight CNN-based face detector suitable for Raspberry Pi deployment. Let the extracted facial image be denoted as I_f.
A CNN-based emotion classifier predicts emotional states:

$$E = \{\text{Neutral,Happy,Anger,Fear}\} \quad (5)$$

Emotion probabilities are computed using the softmax function:

$$P(E_k \mid I_f) = \frac{e^{y_k}}{\sum_{m=1}^{M} e\, y_m} \quad (6)$$

where M represents the number of emotion classes. Facial emotion recognition provides semantic cues related to **human intent**, which cannot be reliably inferred from motion patterns alone.

3.2.4 Activity–Emotion Decision-Level Fusion Strategy

To reduce false alarms and improve reliability, a rule-based decision-level fusion approach combines activity ((A)) and emotion ((E)) predictions:

- Normal activity + Neutral/Happy: Non-Suspicious
- Ambiguous activity + Anger/Fear: Suspicious
- Abnormal activity: Suspicious

This approach uses emotional context to interpret ambiguous activities, improving detection reliability and reducing false positives.

4. RESULTS

a) GUI and Edge Device: Tkinter is used as our GUI in the Raspberry Pi such that we can do live monitoring

control of the camera as well as receive real-time notifications using the buzzer.

b) Facial Emotion Recognition: Haar Cascade is the face detector tool used by our system, followed by aspecial computer software called CNN that identifies emotions on the face. It can identify seven emotionsand represents them through different color representations.

c) Human Activity Detection: MediaPipe Pose generates landmark features for the classification of human

activities as either normal or suspicious with the labeled outputs.

d) Activity–Emotion Fusion & Alerts: There are specific rules set for combining activity and emotion to generate notifications as follows:

- Running + neutral/happy = Normal
- Running + fear/anger = Suspicious
- Fighting + all emotions = Suspicious

Detection of suspicious behavior initiates buzz alerts accompanied by red labels. Screen captures may also be taken to log any incidents.

e) Advantages of the Edge: All computation is done locally on the Raspberry Pi without the need for cloud services and hence gives real-time results at ~6-8 fps. This also minimizes false positives.

Output Visualization

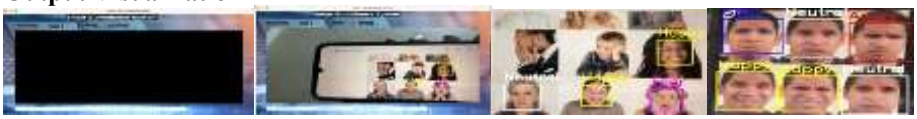


Figure 4.1. GUI Home Screen (with Detection). Figure 4.2 Real-time Emotion Detection Output.



Figure 4.2. Normal Activity Detection (Sitting/Standing)

5. CONCLUSION

This study presents an Edge-IoT surveillance system for real-time human activity, pose, and facial emotion recognition. Using a Raspberry Pi, the system enables fast, privacy-preserving, and independent edge-based decision-making. The proposed framework achieves 96% accuracy, reduces false alarms, and supports efficient surveillance in environments such as schools and public spaces.

6. REFERENCES

[1] M. M. Zaidi et al., "Suspicious human activity recognition from surveillance videos using deep learning," IEEE Access, vol. 12, pp. 105497–105510, 2024, doi: 10.1109/ACCESS.2024.3436653.

[2] M. Biswas et al., "State-of-the-art violence detection techniques: A review," Asian Journal of Computer Science and Information Technology, vol. 13, pp. 29–42, 2022, doi: 10.9734/AJRCOS/2022/v13i130305.

[3] P. Kuppusamy and V. C. Bharathi, "Human abnormal behaviour detection using CNNs in crowded and uncrowded surveillance – A survey," Measurement: Sensors, vol. 24, p. 100510, 2022, doi: 10.1016/j.measen.2022.100510.

[4] M. Khanam, M. Sharif, X. Cheng and S. Kadry, "Suspicious action recognition in surveillance based on handcrafted and deep learning methods: A survey of the state of the art," Computers & Electrical Engineering, vol. 120, p. 109811, 2024.

[5] K. U. Duja, I. A. Khan and M. Alsuhaibani, "Video surveillance anomaly detection: A review on deep learning benchmarks," IEEE Access, vol. 12, pp. 164811–164842, 2024, doi: 10.1109/ACCESS.2024.3491868.

[6] T. Saba et al., "Suspicious activity recognition using proposed deep L4-branched-actionnet with entropy coded ant colony system optimization," IEEE Access, vol. 9, pp. 89181–89197, 2021, doi: 10.1109/ACCESS.2021.3091081.

[7] T. Rehman et al., "Internet-of-things-based suspicious activity recognition using multimodalities of computer vision for smart city security," Security and Communication Networks, vol. 2022, Art. no. 8383461, 2022, doi: 10.1155/2022/8383461.

[8] O. Saidani et al., "An efficient human activity recognition using hybrid features and transformer model," IEEE Access, vol. 11, pp. 101373–101386, 2023, doi: 10.1109/ACCESS.2023.3314492.

[9] N. Dwivedi, D. Singh and D. Kushwaha, "A novel approach for suspicious activity detection with deep learning," Multimedia Tools and Applications, vol. 82, pp. 1–24, 2023, doi: 10.1007/s11042-023-14445-7.

[10] W. Ullah et al., "CNN features with bidirectional LSTM for real-time anomaly detection in surveillance networks," Multimedia Tools and Applications, vol. 80, no. 11, pp. 16979–16995, 2021.

[11] R. Vrskova et al., "A new approach for abnormal human activities recognition based on ConvLSTM architecture," Sensors, vol. 22, no. 8, p. 2946, 2022, doi: 10.3390/s22082946.

[12] M. Q. Gandapur and E. Verdú, "ConvGRU-CNN: Spatiotemporal deep learning for real-world anomaly detection in video surveillance system," Int. J. Interactive Multimedia and Artificial Intelligence, vol. 8, no. 4, p. 88, 2023.

[13] Q. U. A. Arshad et al., "Anomalous situations recognition in surveillance images using deep learning," Computers, Materials & Continua, vol. 76, no. 1, pp. 1103–1125, 2023.

[14] R. Rajeswari, "Anomalous human activity recognition from video sequences using BRISK features and convolutional neural networks," Galaxy International Interdisciplinary Research Journal, vol. 10, no. 2, pp. 269–280, 2022.U. Khan, S. Afzal and J.

[15] W. Lee, "Human activity recognition via hybrid deep learning-based model," Sensors, vol. 22, no. 1, p. 323, 2022, doi: 10.3390/s22010323.

[16] R. R. Subramanian and V. Vasudevan, "A deep genetic algorithm for human activity recognition leveraging fog computing frameworks," Journal of Visual Communication and Image Representation, vol. 77, p. 103132, 2021, doi: 10.1016/j.jvcir.2021.103132.

[17] Z. Xiao et al., "A federated learning system with enhanced feature extraction for human activity recognition," Knowledge-Based Systems, vol. 229, p. 107338, 2021, doi: 10.1016/j.knosys.2021.107338.

[18] Jia et al., "Multi-frequency and multi-domain human activity recognition based on SFCW radar using deep learning," Neurocomputing, vol. 444, pp. 274–287, 2021, doi: 10.1016/j.neucom.2020.07.136.

Biographies



Nidhi P. Nandankar received the bachelor's degree in information technology from Savitribai Phule Pune University in 2022, and the master's degree in computer engineering from Savitribai Phule Pune University in 2026. She is currently working as an Assistant Professor in the Department of Computer Engineering at Modern Education Society's Wadia College of Engineering, Pune- 411001. Her research interests include Internet of Things , Machine Learning.



Dr. S.R. Khonde is an Associate Professor in Computer Engineering at M.E.S. Wadia College of Engineering, Pune, with over 20 years of teaching experience. She holds a Ph.D. in Computer Science and Engineering and has published 45+ journal and 60+ conference papers, along with two textbooks. Her research interests include Data Mining, Cybersecurity, Network Security, and Blockchain Technologies.

OmniAI: A Multi-Agent Conversational Assistant for Task Automation

Priyanka Waghmare[1], Satchal Patil[2], Aditya Savale[3], Prafull Satle[4], Krishna Tilwane[5]

[1] Assistant Professor, Artificial Intelligence & Data Science Department, Ajeenkya DY Patil School of Engineering, Pune, India, priyanka.waghmare@dypic.in

[2345] Artificial Intelligence & Data Science Department, Ajeenkya DY Patil School of Engineering, Pune, India, satchalpatil04@gmail.com, adityasavale74@gmail.com, prafullsatle@gmail.com, krishnatilwane33@gmail.com

Abstract.

Business experts spend a good deal of their time on repetitive duties such as composing emails, preparing reports, running queries on structured data and making charts. There are a number of different types of automation tools, but many of them will need you to have programming skills or commands. These requirements restrict the amount of people who can use automated processes. This study proposes the multi-agent conversational process called OmniAI to give users a way to input into natural language commands that will be translated into digital action. The three main components of the OmniAI system are: intent detection, schema-based query construction and specialized agent systems for processing: emails, reports, visualizations and database queries. Experiments show it achieves an effectiveness of greater than 90 percent for carrying out tasks in each of the modules and has an average response time of less than two seconds, allowing useful perform. The suggested solution can be used in enterprise environments as well as educational institutions to reduce repetitive manual work and enhance efficiencies.

Keywords. Conversational AI; Natural Language Processing; Task Automation; Text-to-Sql; Multi-agent systems

1. INTRODUCTION

Professionals increasingly rely on digital tools to perform routine tasks such as writing emails, generating reports, querying structured data, and creating charts. These repetitive activities often require switching between multiple applications, increasing cognitive load and reducing productivity. Although automation technologies can streamline such workflows, many are designed primarily for technical users and require scripting, predefined commands, or familiarity with APIs. This limits their accessibility to non-technical users. Recent advances in natural language processing (NLP) and large language models (LLMs) provide new opportunities for human-computer interaction. By enabling computers to understand natural-language instructions, these technologies can automate complex digital workflows without requiring users to possess extensive technical knowledge.

2. RELATED WORK

Cognitive automation is moving towards more intelligent and adaptive enterprise systems, with a transitional phase between traditional Robotic Process Automation (RPA) [7] and cognitive automation. In recent research, tasks are delegated to autonomous agents in structured environments, known as agentic information systems [6]. The modern solutions use multi-agent orchestration frameworks which can facilitate coordination of special agents to perform complex tasks [1, 5]. These systems embed domain-specific functionality, such as using large language models to automate tasks [3] and multi-agent systems for data visualization [2]. Also, Text-to-SQL systems can benefit from the feedback obtained during execution to enhance queries precision [2]. The performance of the system is measured by accuracy and latency using a benchmarking system like TaskBench [4].

Existing traditional BI systems rely heavily on ETL-based data access, specialized interfaces, cloud-dependent infrastructure, and manual dashboard creation, limiting usability, privacy, and scalability. A gap remains in developing an integrated, locally deployable system that combines natural-language interaction, direct database access, feedback-driven accuracy, automated visualization, and collaborative reporting within a unified conversational framework.

3. SYSTEM ARCHITECTURE

OmniAI is organised in a five-layer hierarchical architecture, where each layer performs a specific part of the workflow execution pipeline. The layers are intended to be used in a sequential manner, where the output of one layer becomes the input to the other. A summary of all the layers and what they are composed of is given in table.

Table 1. OmniAI Architectural Layers and Component Responsibilities

Layer	**Components**	**Function**
User Interface	Chat, Voice, Dashboard	Accepts queries and displays results
Reasoning	NLP, Intent Router, Model Selector	Understands intent and selects models
Agents	SQL, Report, Email, Scheduling	Executes specific tasks
Data Processing	Database, Pandas, Plotly	Processes data and generates insights
Learning	Feedback Logger, Retraining	Improves performance from feedback

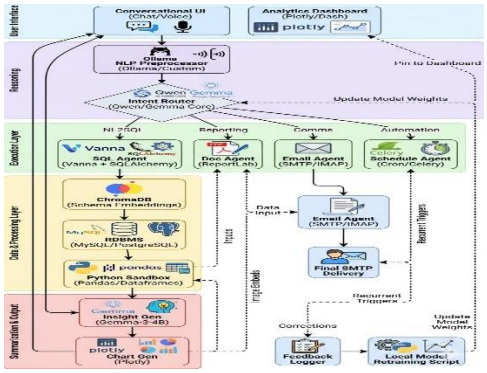


Figure 3.1. System Architecture Diagram

OmniAI processes natural-language queries through intent detection and specialized agents for SQL, reporting, and visualization, delivering results as charts, reports, or emails. Its key contributions include on-premises deployment, Qwen-Gemma multi-model routing, schema-aware NL2SQL with vector retrieval, and a unified conversational pipeline. Compared with LangChain and AutoGen, OmniAI achieves 92.5% query accuracy and reduces execution latency by 20–30% through specialized agent routing and schema-aware processing.

6. METHODOLOGY

It is a system with an agent, called a router, which is central and has knowledge about the user's intentions, and which decides which of the agents to send the task to. The NL2SQL agent converts your queries into SQL, the visualization agent generates graphs, the report agent generates structured output, and the email agent automates email communication.

7. RESULTS

The overall accuracy and success rates for the various tasks are high, in the range of 90% or more, and the reliability and performance of the system are consistently high. Automated emails achieve a 98.7% success rate. The system automates tasks through a natural-language interface with high accuracy and low response time. Its modular design provides flexibility, but performance depends on language-model accuracy and limited data testing. Future work will enhance contextual understanding and extend system capabilities.

Figure 7.1. Text to SQL Conversion

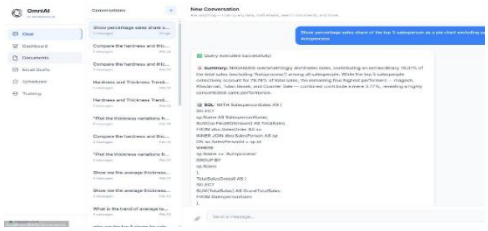
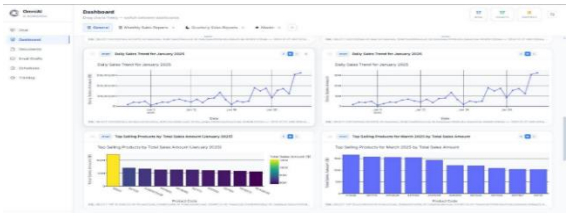


Figure 7.2. Dashboard Data Visualization



9. CONCLUSION

It proposes a multi-agent conversational system to automate digital tasks without the need of technical skills. Such a system has an accuracy of more than 90% in modules that respond within 2 seconds. Future work will focus on scalability, context and enterprise system integration.

10. REFERENCES

[1] J. Hughes et al., "The Orchestration of Multi-Agent Systems: Architectures, Protocols, and Enterprise Adoption," arXiv preprint arXiv:2601.13671, 2026. https://arxiv.org/abs/2601.13671

[2] S. Fernando et al., "VizGen: Data Exploration and Visualization from Natural Language via a Multi-Agent AI Architecture," arXiv preprint arXiv:2509.22218, 2025. https://arxiv.org/abs/2509.22218

[3] R. Novelo, R. R. Silva, and J. Bernardino, "A Literature Review of Personalized Large Language Models for Email Generation and Automation," Future Internet, vol. 17, no. 12, p. 536, 2025. https://doi.org/10.3390/fi17120536

[4] Y. Shen et al., "TaskBench: Benchmarking Large Language Models for Task Automation," arXiv preprint arXiv:2311.18760, 2024. https://arxiv.org/abs/2311.18760

[5] W. Zhou et al., "Agents: An Open-source Framework for Autonomous Language Agents," arXiv preprint arXiv:2309.07870, 2023. https://arxiv.org/abs/2309.07870

[6] C. Engel, P. Ebel, and J. M. Leimeister, "Cognitive Automation," Electronic Markets, vol. 32, no. 1, pp. 339-363, 2022. https://doi.org/10.1007/s12525-021-00519-7

[7] P. Hofmann, C. Samp, and N. Urbach, "Robotic Process Automation," Electronic Markets, vol. 30, no. 1, pp. 99–106, 2020. https://doi.org/10.1007/s12525-019-00365-8

[8] H. Ye et al., "Cognitive Automation: A Survey on the Intersection of AI and Process Automation," ACM Computing Surveys, vol. 56, no. 3, pp. 1–38, 2023 https://doi.org/10.1145/3610214

Biographies



Prof. Priyanka Waghmare is an Assistant Professor in the Department of Artificial Intelligence and Data Science Engineering at Ajeenkya DY Patil School of Engineering Pune. With nearly a decade of experience in academia, Prof. Priyanka has been instrumental in shaping the AI and Data Science curriculum and fostering a research-driven environment at ADYPOSE, Pune. Prof Priyanka is currently pursuing a Ph.D. in Neural Networks.



Satchal Patil is currently pursuing a Bachelor of Engineering in Artificial Intelligence and Data Science at Ajeenkya DY Patil School of Engineering, Lohegaon, Pune. His areas of interest include artificial intelligence, machine learning, and data science.



Aditya Savale is currently pursuing a Bachelor of Engineering in Artificial Intelligence and Data Science at Ajeenkya DY Patil School of Engineering, Lohegaon, Pune. His areas of interest include artificial intelligence, machine learning, and data science.



Prafull Satle is currently pursuing a Bachelor of Engineering in Artificial Intelligence and Data Science at Ajeenkya DY Patil School of Engineering, Lohegaon, Pune. His areas of interest include artificial intelligence, machine learning, and data science.



Krishna Tilwane is currently pursuing a Bachelor of Engineering in Artificial Intelligence and Data Science at Ajeenkya DY Patil School of Engineering, Lohegaon, Pune. His areas of interest include artificial intelligence, machine learning, and data science.

An Integrated Intelligent System for Transformer Manufacturing

V.Gedam, S. Gimhavanekar, S. Ghumre, R. Gaikwad, V. Gaikwad

Assistant Professor, Artificial Intelligence & Data Science Department, Ajeenkya DY Patil School of Engineering, Pune, India, vishakhagedam@dypic.in, shreyasgimhavanekar582003@gmail.com, ghumresangram@gmail.com, ruturajrsg029@gmail.com, vaibhavraje5044x@gmail.com,

Abstract.

In this paper, the development of an integrated intelligent system in the transformer manufacturing is introduced, which is based on the three modules of forecast, demand modelling and conversational AI and predictive maintenance. The solution to inefficiencies arising from isolated solutions. A Random Forest model predicts failures and a forecasting model optimizes inventory. Results indicate lowered downtime and more efficiency. It is suitable for usage in industrial automation and smart manufacturing. Keywords: predictive maintenance; machine learning; transformer manufacturing; automation; AI

Keywords. Predictive maintenance, machine learning, transformer manufacturing, demand forecasting, conversational AI, industrial automation, artificial intelligence.

1. INTRODUCTION

Transformer manufacturing faces challenges related to equipment failures, inventory management, and customer service. These issues are often handled using separate systems, resulting in data silos and inefficient decision-making. This paper proposes an integrated intelligent system that combines predictive maintenance, demand forecasting, and conversational AI. The system aims to reduce downtime, improve inventory planning, and provide faster responses to operational and customer queries.

2. RELATED WORK

Machine learning techniques such as Random Forest, neural networks, and deep learning have been widely used for predictive maintenance and equipment health monitoring. Industry 4.0 and IoT technologies have also enabled large-scale data collection and intelligent decision-making. Similarly, conversational AI has been increasingly used for automated customer support. However, most existing solutions focus on these applications individually rather than integrating predictive maintenance, demand forecasting, and conversational AI into a single system.

3. PROPOSED INTEGRATED SYSTEM AND METHODOLOGY

The proposed system integrates three intelligent modules: predictive maintenance, demand forecasting, and conversational AI. Operational data collected from a transformer production facility over 18 months was used for evaluation, including telemetry, maintenance, inventory, and customer data. The predictive maintenance module uses machine learning models to predict equipment failures, while the demand forecasting module supports inventory planning. The conversational AI module provides contextually relevant responses to operational and customer queries. By integrating these modules, the system enables better data sharing and coordinated decision-making across manufacturing operations.

4. RESULTS AND DISCUSSION

4.1. Predictive Maintenance Results

The predictive maintenance models were evaluated using Accuracy, AUROC, Precision, Recall, and F2-score. Among the evaluated models, Random Forest achieved the best overall performance, with an accuracy of 91.3%, AUROC of 0.89, and F2-score of 0.85. Its recall of 0.86 indicates effective failure detection.

Table 1. Predictive Maintenance Performance

Model	**Accuracy**	**AUROC**	**Precision**	**Recall**	**F2-score**
Logistic Regression	0.823	0.81	0.72	0.69	0.70
Random Forest	0.913	0.89	0.84	0.86	0.85
XGBoost	0.901	0.87	0.82	0.83	0.83
LSTM	0.887	0.85	0.79	0.84	0.82

The combined use of telemetry, temporal, and historical features achieved better performance than individual feature groups.

4.2. Demand Forecasting and Operational Results

The demand forecasting model achieved an sMAPE of 8.7% and a MASE of 0.83. Higher forecasting accuracy was observed for high-volume SKUs compared with low-volume products. The integrated system also showed significant operational improvements, including reduced downtime, improved inventory turnover, and faster query resolution.

Table 2. Operational Performance Comparison

Metric	**Baseline**	**With System**	**Improvement**
Downtime (hrs/month)	47.3	31.2	↓34%

Inventory Turnover	4.2	5.4	↑28.6%
Query Resolution	8.4	2.8	↓66.7%

These results demonstrate that integrating predictive maintenance, demand forecasting, and conversational AI can improve coordination and operational efficiency in transformer manufacturing.

5. CONCLUSION

This study developed an integrated AI system for transformer manufacturing combining predictive maintenance, demand forecasting, and conversational AI. Random Forest achieved 91.3% accuracy, while forecasting achieved 8.7% sMAPE. The system reduced downtime by 34%, improved inventory turnover by 28.6%, and reduced query resolution time by 66.7%. Future work will evaluate the system across multiple facilities and longer periods.

6. REFERENCES

[[1] A. Sharma, "Machine learning techniques for maintenance systems of transformers in industries," IEEE Transactions on Industrial Informatics, vol. 19, no. 4, pp. 2751–2763, Apr. 2023, doi: 10.1109/TII.2023.1004582.

[2] M. K. Gupta and L. S. Chauhan, "Design and deployment of an AI-based predictive maintenance and monitoring system in manufacturing industries," Journal of Intelligent Manufacturing Systems, vol. 35, no. 2, pp. 412–426, Feb. 2024, doi: 10.1016/j.jims.2024.02.011.

[3] D. Patel and N. Joshi, "Machine learning-based fault detection and predictive analytics of smart industrial systems," Procedia Computer Science, vol. 226, pp. 1289 1301, 2023, doi: 10.1016/j.procs.2023.03.162.

[4] H. Zhang, Y. Wang, and K. Lee, "Deep learning models for real-time predictive maintenance in power transformer monitoring," IEEE Access, vol. 11, pp. 98412–98425, Oct. 2023, doi: 10.1109/ACCESS.2023.3298456.

[5] R. S. Bhosale, P. R. Deshmukh, and T. M. Gaikwad, "Workflow automation and cloud integration in industrial asset management," International Journal of Emerging Technologies in Engineering Research (IJETER), vol. 12, no. 1, pp. 88–95, Jan. 2024.

[6] L. Fernandes, S. R. Pillai, and M. George, "AI-based chatbots for industrial customer care using NLP and intent recognition," IEEE Internet Computing, vol. 28, no. 3, pp. 34–42, May 2024, doi: 10.1109/MIC.2024.1031234.

[7] N. K. Singh and R. Agarwal, "Material demand forecasting in manufacturing using hybrid ARIMA-XGBoost models," Journal of Manufacturing Systems, vol. 77, pp. 232–244, Dec. 2023, doi: 10.1016/j.jmsy.2023.12.008.

[8] M. Das and S. Kumar, "A secure web-based industrial automation portal using Flask and AWS Lambda," International Journal of Computer Applications, vol. 186, no. 45, pp. 50–56, 2024.

[9] P. Singh and R. Tiwari, "Predictive analytics in transformer maintenance: A case study on fault forecasting using data analysis," Energy Reports, vol. 12, pp. 1903–1914, Aug. 2024, doi: 10.1016/j.egyr.2024.08.021.

[10] T. P. Carvalho et al., “A systematic literature review of machine learning methods applied to predictive maintenance,” Computers and Industrial Engineering, vol. 137, p. 106024, 2019, doi: 10.1016/j.cie.2019.106024.
[11] G. A. Susto et al., “Machine learning for predictive maintenance: A multiple classifier approach,” IEEE Transactions on Industrial Informatics, vol. 11, no. 3, pp. 812–820, Jun. 2015, doi: 10.1109/TII.2014.2349359.
[12] S. Khan and T. Yairi, “A review on the use of deep learning in system health management,” Mechanical Systems and Signal Processing, vol. 107, pp. 241–265, May 2018, doi: 10.1016/j.ymssp.2017.11.024.
[13] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” Mechanical Systems and Signal Processing, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/j.ymssp.2005.09.012.
[14] J. Lee, B. Bagheri, and H.-A. Kao, “A cyber-physical systems architecture for Industry 4.0-based manufacturing systems,” Manufacturing Letters, vol. 3, pp. 18–23, Jan. 2015, doi: 10.1016/j.mfglet.2014.12.001.
[15] J. Lee, H.-A. Kao, and S. Yang, “Service innovation and smart analytics in Industry 4.0 and big data environment,” Procedia CIRP, vol. 16, pp. 3–8, 2014, doi: 10.1016/j.procir.2014.02.001.
[16] “An Introduction to Predictive Maintenance,” 2nd ed., Burlington, MA, USA: Butterworth-Heinemann, 2002.
[17] D. Mourtzis, E. Vlachou, and N. Milas, “Industrial big data as a result of IoT adoption in manufacturing,” Procedia CIRP, vol. 55, pp. 10.1016/j.procir.2016.08.041. 290–295, 2016, doi:
[18] E. Adamopoulou and L. Moussiades, “Chatbots: History, technology, and applications,” Machine Learning with Applications, vol. 2, p. 100006, 2020, doi: 10.1016/j.mlwa.2020.100006.
[19] A. Bousdekis, D. Apostolou, and G. Mentzas, “Predictive maintenance in the 4th industrial revolution: A literature review,” in Proc. Int. Conf. Information, Intelligence, Systems and Applications (IISA), Patras, Greece, 2019, pp. 1–8, doi: 10.1109/IISA.2019.8900668.

Biographies



Sangram Ghumre is affiliated with the Ajeenkya DY Patil School of Engineering, Pune, India. His research interests include Artificial Intelligence, Machine Learning, Generative AI, Retrieval-Augmented Generation (RAG), Natural Language Processing, and Intelligent Automation.



Ruturaj Gaikwad is affiliated with the Ajeenkya DY Patil School of Engineering, Pune, India. His research interests include Artificial Intelligence, Machine Learning, Generative AI, Retrieval-Augmented Generation (RAG), Natural Language Processing, and Intelligent Automation.



Shreyas Gimhavanekar is affiliated with the Ajeenkya DY Patil School of Engineering, Pune, India. His research interests include Artificial Intelligence, Machine Learning, Generative AI, Retrieval-Augmented Generation (RAG), Natural Language Processing, and Intelligent Automation.



Vaibhav Gaikwad is affiliated with the Ajeenkya DY Patil School of Engineering, Pune, India. His research interests include Artificial Intelligence, Machine Learning, Generative AI, Retrieval-Augmented Generation (RAG), Natural Language Processing, and Intelligent Automation.



Prof Vishakha Gedam is currently working in the Department of Artificial Intelligence and Data Science (AIDS), Ajeenkya DY Patil School of Engineering, Pune, India, with nearly 10 years of academic experience in engineering education. Her professional experience includes teaching, student mentoring, academic activities, and the application of emerging technologies in education. She is currently pursuing a Ph.D. at Symbiosis Institute of Technology, Pune, India.

The Experiential Roadmap to Launching Your Venture

Disha S. Wankhede, Parth Beldar, Baswadeep Bembarem, Aatish Bagal, Savya Bhamare

Vishwakarma Institute of Technology, Pune, disha.wankhede@vit.edu, parthbeldar85@gmail.com, brvbaswadeep@gmail.com, aatishbagal135@gmail.com, bhamaresavya609@gmail.com

Abstract.

Entrepreneurship is a key driver of innovation and employment, yet conventional education often lacks practical startup experience. This paper presents Scalvia, a simulation-based experiential learning platform that enables students and aspiring entrepreneurs to create and manage virtual startups. Users make strategic decisions involving funding, operations, marketing, competition, and business scaling. Gamification features, including levels, rewards, performance scores, and feedback analytics, promote engagement and continuous learning. By combining realistic simulation with interactive learning, Scalvia bridges the gap between theory and practice while developing entrepreneurial resilience, risk awareness, decision-making skills, and confidence. The platform offers an innovative approach to entrepreneurship education.

Keywords: Entrepreneurship Education; Experiential Learning; Startup Simulation; Venture Development; Gamification.

1. INTRODUCTION

Entrepreneurship drives global innovation, employment, and sustainable economic development. However, traditional entrepreneurship education often emphasizes theory and provides limited exposure to real startup challenges involving funding, resources, marketing, customer retention, and scaling. This paper presents Scalvia, a simulation-based platform that enables learners to build and manage virtual startups in a risk-free environment. Users make decisions related to funding, marketing, operations, and scaling while experiencing realistic outcomes. Gamification enhances engagement and motivation. Scalvia helps learners develop confidence, critical thinking, problem-solving, decision-making, and practical entrepreneurial skills while safely learning from both success and failure [1-5].

2. RELATED WORK

Education is shifting from traditional lectures toward interactive and experiential learning. In entrepreneurship education, simulation-based learning enables students to apply theoretical concepts in realistic environments while developing critical thinking, collaboration, and decision-making skills [3-9]. Gamification further enhances motivation, engagement, creativity, and confidence. Research indicates that combining simulation and game-based learning helps students understand business dynamics, practice decision-making, and develop entrepreneurial competencies through continuous interaction with realistic scenarios [10-17]. These approaches bridge theory and practice by allowing learners to experiment with risks and failures safely. Building on these insights, Scalvia provides an

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

immersive startup learning experience that integrates simulation-based learning with gamification.

Table 1. Compares this work with the related work or previous research by other researchers

Ref.	Simulation Based Learning	Gamification	Decision Making Skills	Critical Thinking	Safe Risk-Free Environment	Read World Business Environment
[1]	Yes	No	Yes	Yes	Yes	Yes
[2]	No	Yes	No	Yes	Yes	No
[3]	No	Yes	Yes	No	Yes	Yes
[4]	Yes	No	Yes	No	No	Yes
[5]	No	Yes	Yes	No	Yes	No
[6]	Yes	No	Yes	No	No	Yes
[7]	Yes	No	Yes	No	No	Yes
[8]	Yes	No	Yes	No	No	Yes
[9]	Yes	No	Yes	No	No	Yes
[10]	No	Yes	Yes	No	Yes	No
This work	Yes	Yes	Yes	Yes	Yes	Yes

3. METHOD, EXPERIMENTS AND RESULTS

Simulation-based learning enables students to apply theoretical knowledge in realistic entrepreneurial settings, while gamification improves motivation, engagement, creativity, and confidence. Scalvia combines both approaches to support startup creation and management. The study defines the scoring models and algorithms used to simulate startup decisions and outcomes using deterministic and probabilistic methods, following established startup ecosystem modeling approaches [17].

The Minimum Viable Product (MVP) viability score evaluates the initial quality of a startup by considering founder capabilities, team effectiveness, financial conditions, and operational risks. The final score is constrained to a practical range to avoid extreme values:

$$S_{mvp} = clamp(S_{raw}, 6{,}15) \quad (1)$$

where the raw score combines founder skills, team performance, financial support, team synergy, and risk penalties. Founder competency is calculated from technical and business skills, with greater emphasis on coding and design:

$$S_{founder} = 0.3 \cdot k_{code} + 0.25 \cdot k_{design} + 0.15 \cdot k_{strat} + 0.1 \cdot \left(k_{data} + k_{fin} + k_{ops}\right) \quad (2)$$

Team effectiveness considers employee skills, experience, and morale while accounting for diminishing returns as team size increases:

$$S_{team} = \sqrt{10 \cdot \sum_{i=1}^{n} \left(\underline{s_i} \cdot (0.6 + 0.04 \cdot e_i) \cdot \frac{m_i}{10}\right)} \quad (3)$$

After launch, product quality, team performance, marketing, and market uncertainty influence startup growth. User growth is modeled using the existing user base, marketing effort, and product quality:

$$\Delta U_{growth} = (0.1 \cdot U_t, 100) \cdot M(S_{strat}) \cdot \frac{Q}{10} \cdot (0.8 + \epsilon) \quad (4)$$

Other simulation components, including product iteration, retention, technical debt, investment pitching, and crowdfunding success, are modeled using weighted factors, diminishing-return functions, penalties, and stochastic adjustments. These mechanisms allow Scalvia to represent realistic startup uncertainty without requiring numerous separate equations. Initial testing indicated that the simulation and gamification features improved engagement, motivation, decision-making, and confidence. Participants experienced funding, resource management, competition, and scaling scenarios while learning from simulated outcomes. However, further evaluation is required to assess scalability, realism, and long-term educational effectiveness. Overall, Scalvia provides a safe and interactive environment for developing entrepreneurial decision-making, critical thinking, risk awareness, resilience, and confidence.

4. PHOTOGRAPHS AND FIGURES

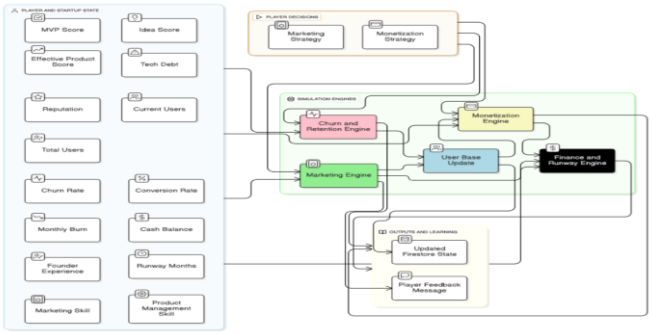


Figure 4.1. Startup Simulation Engine Architecture: Player Decisions, Metrics, and System Flow.

5. CONCLUSIONS

Scalvia provides an immersive and practical platform for entrepreneurship education by simulating the complete startup journey. It bridges theoretical learning with real-world decision-making while developing critical thinking, problem-solving, resilience, and leadership skills. Although scalability, regular updates, and gamification remain challenges, future work will focus on empirical validation, larger user groups, and improved realism. Overall, Scalvia offers an effective experiential learning environment for developing next-generation entrepreneurs.

6. REFERENCES

[1]. A. T. Rosário and R. A. Raimundo, "Enhancing business higher education through simulation-based learning, problem-based learning and challenge-based learning: A systematic bibliographic literature review," Preprints, 2024. [Online]. Available: https://doi.org/10.20944/preprints202407.0747.v1

[2]. F. Keshmiri, N. Zeraati, S. Jalali, and M. Zeraati, "The effect of gamification in entrepreneurship and business education," Frontiers in Education, vol. 10, art. 144, 2025. [Online]. Available: https://doi.org/10.3389/feduc.2025.11974179

[3]. A. Isabelle, "Gamification of entrepreneurship education," Decision Sciences Journal of Innovative Education, vol. 18, no. 2, pp. 203–223, 2020. [Online]. Available: https://doi.org/10.1111/dsji.12207

[4]. I. Lee and T. K. Yu, "Game-based learning enhances business decision-making learning for on-the-job MBA students: A case study of dynamic systems-thinking course," The International Journal of Management Education, vol. 23, art. 101140, 2025. [Online]. Available: https://doi.org/10.1016/j.ijme.2025.101140

[5]. M. Lynch, E. Kubberød, N. Sanne, and A. H. F. Josendal, "Deliberate Practice Through the Gamification of Entrepreneurship Education," Entrepreneurship Education and Pedagogy, vol. 8, no. 3, pp. 485–510, 2025. [Online]. Available: https://doi.org/10.1177/25151274241309940

[6]. S. A. M. Gilani, N. Yasin, and H. Hafeez, "Simulation-based learning in business and entrepreneurship in higher education: A review of the games available," Education and Information Technologies, vol. 27, no. 4, pp. 4899–4921, 2022. [Online]. Available: https://doi.org/10.1007/s10639-021-10713-5

[7]. C. Chatpinyakoop, P. Hallinger, and P. Showanasai, "Assessing the effects of online simulation-based learning on skills in managing change for corporate sustainability," The International Journal of Management Education, vol. 22, art. 100960, 2024. [Online]. Available: https://doi.org/10.1016/j.ijme.2024.100960

[8]. T. Takemoto and H. Oe, "Entrepreneurship education at universities: Challenges and future perspectives on online game implementation," Entrepreneurship Education, vol. 4, pp. 149–167, 2021. [Online]. Available: https://doi.org/10.1007/s41959-021-00054-9

[9]. H. Sziegat, "Virtual simulation games in entrepreneurship education: Status quo and prospects," in Proc. 18th European Conf. Games Based Learning (ECGBL 2024), 2024, pp. 642–650. [Online]. Available: https://doi.org/10.34190/ecgbl.18.1.2581

[10]. D. Seguí-Mas, E. Seguí-Mas, and G. Tormo-Carbó, "Gamification in entrepreneurship education: A systematic literature review and research agenda," SSRN Electronic Journal, 2024. [Online]. Available: https://doi.org/10.2139/ssrn.4776641

[11]. L. Neriz, "Simulation-based training as a teaching and learning tool for management education," Innovations in Education and Teaching International, vol. 57, no. 4, pp. 432–443, 2020. [Online]. Available: https://doi.org/10.1080/14703297.2019.1631874

[12]. A. Kara and L. M. Russell, "How to embed gamification into entrepreneurship courses to enhance student engagement," Small Business Institute Journal, vol. 20, no. 2, pp. 28–31, 2024. [Online]. Available: https://sbij.org/index.php/SBIJ/article/view/427

[13]. AACSB Insights, "Simulations spark engagement and real learning," AACSB International, Sep. 2022. [Online]. Available: https://www.aacsb.edu/insights/articles/2022/09/simulations-spark-engagement-and-real-learning

[14]. L. Chen, D. Ifenthaler, J. Y. Yau, and W. Sun, "Artificial intelligence in entrepreneurship education: a scoping review," Education + Training, vol. 66, no. 6, pp. 589–608, 2024. [Online]. Available: https://doi.org/10.1108/ET-05-2023-0169

[15]. C. Moldez, M. A. Crisanto, G. R. Cerdeña, S. Maranan, and R. Figueroa, "Innovation in education: Developing and assessing gamification in massive open online courses," arXiv:2409.03309 [cs.CY], 2024. [Online]. Available: https://arxiv.org/abs/2409.03309 s

[16]. J. Hamari and B. K. Daneva, "Game design patterns for learning: A review of how gamification elements influence user behavior," Computers in Human Behavior, vol. 141, art. 107628, 2023. [Online]. Available: https://doi.org/10.1016/j.chb.2022.107628

[17]. A. Silva and T. B. Murthy, "Mathematical modeling of startup ecosystems for digital twins and simulations," Journal of Business Venturing Insights, vol. 21, art. e00452, 2024. [Online]. Available: https://doi.org/10.1016/j.jbvi.2023.e00452

Biographies



Dr. Disha Sushant Wankhede is a Postdoctoral Fellow at Universiti Putra Malaysia, specializing in Artificial Intelligence, Deep Learning, Image Processing, and Healthcare. She has published 50+ Scopus-indexed papers, holds 20+ patents, and has contributed to numerous research and industry projects. She also serves as a reviewer, session chair, expert speaker, and professional member of ACM, ISTE, and CSI.



Savya Bhamare, Computer Engineering Student at Vishwakarma Institute of Technology, Pune



Baswadeep Bembare, Computer Engineering Student at Vishwakarma Institute of Technology, Pune



Aatish Bagal, Computer Engineering Student at Vishwakarma Institute of Technology, Pune



Parth Beldar, Computer Engineering Student at Vishwakarma Institute of Technology, Pune

Hospital management through AI: smart health assistant app

Prof.Priyanka Waghmare[1], Suvidya Sonawane[2], Sonal Shinde[3], Suraj Dombe[4], Parth Nalawade[5]

[1] Assistant Professor, Artificial Intelligence & Data Science Department, Ajeenkya DY Patil School of Engineering, Pune, India.priyanka.waghmare@dypic.in

[2345] Artificial Intelligence & Data Science Department, Ajeenkya DY Patil School of Engineering, Pune, India.suvidyasonawane15@gmail.com, sonal2004shinde@gmail.com, suraj25dombe@gmail.com, parthnalawade1817@gmail.com

ABSTRACT:

Hospital tasks are often performed manually, making them inefficient and time-consuming. The proposed smart health assistant uses artificial intelligence to improve hospital services through a role-based mobile application for administrators, doctors, and patients. It supports symptom analysis, appointment scheduling, medical record management, health tracking, diet planning, pregnancy care, and visual health statistics. Testing demonstrated improved coordination, efficient appointment management, and reduced dependence on manual operations.

Keywords: Hospital Management, Artificial Intelligence, Healthcare Application, Role Based Access, and Smart Health System.

1. INTRODUCTION

Hospitals increasingly adopt digital solutions to improve daily operations, yet many still rely on traditional methods for appointments, patient records, and communication, causing delays, workload, and potential errors that may affect care quality [1-5]. Artificial intelligence (AI) offers opportunities to automate routine tasks, improve information management, support data analysis, and assist healthcare professionals in decision-making. This project proposes a smart health assistant application with role-based access for administrators, doctors, and patients. The system supports appointment booking, electronic health record management, basic symptom assistance, health monitoring, and supervised patient guidance. The goal is to develop a practical, secure, and user-friendly platform for efficient digital healthcare management.

2. RELATED WORK

The healthcare sector is rapidly adopting digital technologies to improve efficiency and quality of care. Artificial intelligence (AI) can support clinical decision-making, reduce routine workload, and analyze patient data to retrieve essential information for diagnosis [1]. Cloud computing and Internet of Things (IoT) technologies further support smart hospital development

by enabling real-time information sharing among healthcare units [2]. Electronic health records (EHRs) have also become essential in modern hospitals, with machine learning increasingly applied to EHR data to identify clinically relevant patterns and relationships [3]. In addition, digital symptom checkers and online health assistants can provide basic guidance based on user inputs, improving access to health information and preliminary support [9]. These developments demonstrate the potential of integrated digital healthcare systems to improve hospital management and patient services. Existing studies [6-9] rarely integrate doctor care, appointment management, medical records, and health monitoring within one platform. The proposed system addresses this gap by combining AI with role-based access for secure hospital management. It integrates appointments, records, symptom guidance, health visualization, and doctor-supervised care into a unified mobile healthcare platform..

4. SYSTEM ARCHITECTURE

Figure 1 illustrates the architecture of the proposed smart health assistant, designed as a client-server system. Administrators, doctors, and patients access the system through a mobile application connected to a centralized database. Role-based access controls information and functions according to user responsibilities. The application processes requests centrally and provides relevant services, including appointment management, patient records, prescriptions, and system monitoring.

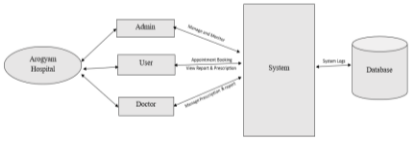


Figure 1. System Architecture Diagram

5. METHODOLOGY

The system was developed to address hospital management challenges, including appointment procedures and disorganized patient records. Role-based access was implemented for administrators, doctors, and patients, with functions tailored to their responsibilities. Developed using Flutter, the application includes authentication, appointment scheduling, medical records, chatbot, health tracking, diet planning, and pregnancy care modules. An AI-based chatbot provides basic health guidance, while appointment management supports scheduling and token generation. Testing demonstrated reduced manual effort, smooth task execution, and improved hospital management efficiency.

6. RESULTS

The developed system was evaluated using four user roles: Admin, Senior Doctor, Junior Doctor, and Patient. Each role was tested with its assigned features to verify module functionality. The system efficiently supported appointment booking, medical record management, and overall interaction through role-specific dashboards. Testing also confirmed effective data security and

privacy. The centralized platform simplified user tasks by providing essential functions in one system. Interfaces for each user role are presented in Figures 2.1–2.4.

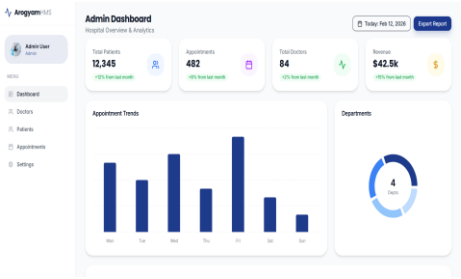


Figure 2.1. Admin Dashboard

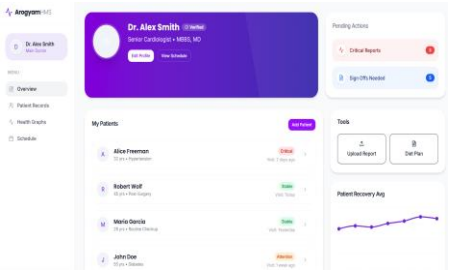


Figure 2.2. Senior Doctor Dashboard



Figure 2.3. Junior Doctor Dashboard

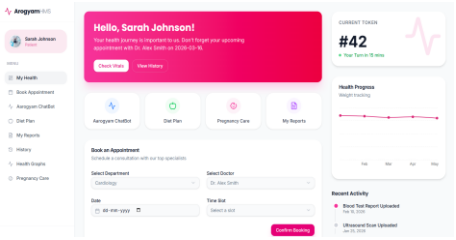


Figure 2.4. Patient Dashboard

8. CONCLUSION

The proposed smart health assistant simplifies digital hospital management by integrating role-based access for administrators, doctors, and patients with appointment scheduling, medical record management, and AI-based assistance. Testing showed improved organization, reduced manual tasks, faster information access, and effective collaboration among user roles. Health tracking and assistant guidance also supported patient engagement. Future development should incorporate real-time database connectivity, advanced AI capabilities, and improved scalability to support larger healthcare environments and enable practical hospital-wide deployment.

9. REFERENCES

[1] H. Kwon, S. Lee, and J. Park, “Review of smart hospital services in real healthcare: IoT, analytics and cloud integration”, Healthcare Informatics Research, vol. 28, no. 2, pp. 123–135, 2022. https://pubmed.ncbi.nlm.nih.gov/35172086/

[2] C. N. Vorisek, D. P. Jones, and M. Li, “Fast healthcare interoperability resources (FHIR) for health research: Systematic review of implementations and use cases”, PeerJ Computer Science, vol. 8, p. e1274, 2022. https://pubmed.ncbi.nlm.nih.gov/35852842/

[3] A.Sharma, P. K. Singh, P. Nikashina and V. Gavrilenko, “IoT and AI-Based Smart Healthcare Monitoring System”, in Data Science and Artificial Intelligence for Digital Healthcare, pp. 195-209, 2024. https://doi.org/10.1007/978-3-031-56818-3_12

[4] A. Al Kuwaiti et al., “A review of the role of artificial intelligence in healthcare”, Frontiers in Public Health, vol. 11, 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC10301994/
[5] C. S. Kruse and A. Beane, “Health information technology continues to show positive effect on medical outcomes: Systematic review”, JMIR Medical Informatics, vol. 6, no. 3, p. e41, 2018. https://www.jmir.org/2018/2/e41/
[6] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, “Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine”, Database, vol. 2020, p. baaa010, 2020. https://pubmed.ncbi.nlm.nih.gov/32185396/
[7] F. Jiang et al., “Artificial intelligence in healthcare: Past, present and future”, Stroke and Vascular Neurology, vol. 2, no. 4, pp. 230–243, 2017. https://pubmed.ncbi.nlm.nih.gov/29507784/
[8] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, “ Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis”, Journal of Biomedical Informatics, vol. 83, pp. 168–185, 2018. https://pubmed.ncbi.nlm.nih.gov/29989977/
[9] M. Hammoud, S.Douglas, M.Darmach, S.Alawneh. S. Sanyal, Y. Kanbour. “Evaluating the Diagnostic Performance of Symptom Checkers: Clinical Vignette Study”, JMIR AI, 2024. https://ai.jmir.org/2024/1/e46875

Biographies



Prof. Priyanka Waghmare is an Assistant Professor in the Department of Artificial Intelligence and Data Science Engineering at Ajeenkya DY Patil School of Engineering Pune. With nearly a decade of experience in academia, Prof. Priyanka has been instrumental in shaping the AI and Data Science curriculum and fostering a research-driven environment at ADYPOSE, Pune. Prof Priyanka is currently pursuing a Ph.D. in Neural Networks.



Suvidya Sonawane is an ambitious Artificial Intelligence and Data Science graduate (Class of 2026) from the Ajeenkya DY Patil School of Engineering, Pune. Driven by a strong interest in

machine learning, data analytics, and modern computing, Suvidya is eager to apply academic knowledge to real-world challenges.



Sonal Shinde is an ambitious Artificial Intelligence and Data Science graduate (Class of 2026) from the Ajeenkya DY Patil School of Engineering, Pune. Driven by a strong interest in machine learning, data analytics, and modern computing, Suvidya is eager to apply academic knowledge to real-world challenges.



Suraj Dombe is an ambitious Artificial Intelligence and Data Science graduate (Class of 2026) from the Ajeenkya DY Patil School of Engineering, Pune. Driven by a strong interest in machine learning, data analytics, and modern computing, Suvidya is eager to apply academic knowledge to real-world challenges.



Parth Nalawade is an ambitious Artificial Intelligence and Data Science graduate (Class of 2026) from the Ajeenkya DY Patil School of Engineering, Pune. Driven by a strong interest in machine learning, data analytics, and modern computing, Suvidya is eager to apply academic knowledge to real-world challenges.

NeuroFlow IDE: An AI-Assisted Interactive Browser-Based Development Environment

Aniket Atram, Ansh Bagul, Yash Bhongale, Ruchika Birajdar, Shobha Patil

Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India.aniket.atram24@vit.edu, yash.bhongale241@vit.edu, ruchika.birajdar24@vit.edu

Abstract.

Desktop and most cloud-based IDEs usually fail to integrate code writing, execution, debugging, assistance and visualization in unified and learner-friendly environment at single platform, creating a challenge for beginners and developer leading to affect learning and productivity timing. To tackle this problem, this project presents NeuroFlow IDE, a cloud-based integrated development environment that contains features like AI-assistance support for coding, learning support, and real time visualization of code flow of how code is working in real time. It is built on React and Monaco Editor with secured Dockerized backend and dynamically powered on D3.js for Visualization. NeuroFlow simplifies the setup process, promotes quicker debugging, and is interactive, allowing an environment if editing, executing, visualizing, and combining multiple platform features into one. Increases in learning and productivities but also contributes substantially to enhancement of practical skills and technical knowledge among the students and developers.

Keywords. Cloud-based IDE; AI-assisted Coding; Real-time Code Visualization; Assertive Debugging; Intelligent Code Analysis; Software Learning Platform.

1. INTRODUCTION

Integrated Development Environments (IDEs) provide essential tools for coding, compilation, debugging, and version control. While traditional IDEs such as IntelliJ IDEA, Eclipse, and Visual Studio are powerful, their setup can be complex for beginners. Cloud IDEs simplify development but often lack real-time execution visualization and intelligent debugging. To address this gap, NeuroFlow IDE combines AI-assisted debugging, real-time program visualization, and secure Docker-based code execution in a unified web environment, supporting effective learning and productivity.

2. RELATED WORK

Independent studies have focused on AI-assisted programming feedback, visualization-based learning tools, and intelligent debugging systems in the recent past. Large language models that generate hints in the style of an AI tutor are helpful in improving programming guidance, but do not support visualization [1]. Improved compiler error-messages research enhances, yet concentrates on syntax-level cognition, in the performance of learners [2]. Visualization systems based on algorithms and VR learning environments enhance the clarity of the concepts but fail to incorporate AI support [3][4]. Code improvement methods and language model-based build optimization techniques are used to improve efficiency of

compilation and debugging of a program [5][7]. Visual debugging tools enable execution tracing but are separate environments [8]. Nevertheless, there has not been a single platform which combines AI support, visualization, and real-time debugging; NeuroFlow IDE tries to fill this gap.

	AI Assistance	Visualization Support	Real-time Debugging	Unified Platform
[1]	Yes	No	No	No
[2]	Yes	No	No	No
[3]	No	Yes	No	No
[4]	No	Yes	No	No
[7]	Yes	No	No	No
[8]	No	Yes	Yes	No
This work	Yes	Yes	Yes	Yes

Table 1. Comparison between this work with the related work or previous research.

3. METHOD, EXPERIMENTS AND RESULTS

This section explains the design and implementation of proposed NeuroFlow IDE, experimental evaluation and performance measurement of proposed NeuroFlow IDE.

3.1. System Architecture

NeuroFlow IDE is based on three layers with the frontend, backend, and sandboxed execution layers as presented as Fig. 3.1.1. The frontend interface is a coding interface that gives interactive execution of displaying programs in real-time. The backend interface receives the execution requests, does a static analysis, and links to the AI services to assist in contextual debugging. The worker execution layer is safe as it operates the program with safety and creates execution traces which are visualized.

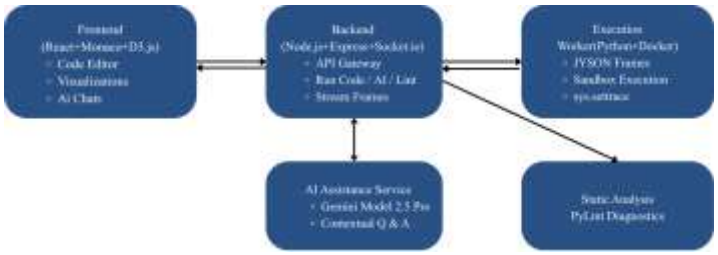


Figure 3.1.1. NeuroFlow System Architecture

3.2. Implementation

The frontend is built with React.js and Monaco Editor in order to create the minimal and fast coding environment resembling the contemporary desktop IDEs. D3.js visualization modules are dynamically used to execute with the help of algorithms like sorting tasks, recursion trees, stacks, queue, and linked lists execution flow. Backend is written in node.js, and it is based on the WebSocket communications between the execution environment and the editor for real-time interaction. Python has the sys.settrace() function which keeps line-

by-line track of execution states, variable assignments, and runtime exceptions, which is used to execute tracing. These traces have been converted into JSON frames and sent to the visualization frontend module. Fig. 3.2.1 represents the workflow of the tracing of the execution and its visualization.



Figure 3.2.1. NeuroFlow System Architecture

A contextual explanation, syntax correction proposal feature, and debugging guidance are available as an integral part of Gemini models with an AI assisting module through which an individual asks the computer questions. The Docker containers, which have restricted CPU, memory, and execution, are used to offer safety in sandboxing of code to avoid unsafe code execution.

3.3. Experimental Setup

This was also evaluated using standard programming tasks such as sorting algorithms, tracing of recursion, and simple data-structure operations. The tested system had 14 participants who were novice and intermediate programmers. Measures of evaluation were the accuracy of executions, synchronization of display, the reliability of AI-responses, and efficiency of system debugging, and stability through variability of browsers and devices

3.4. Result and Performance Evaluation

The results proved in improvement of about 65-75% in conceptual understanding of algorithm execution compared to the traditional IDE environments e.g. VS Code and Code Blocks. Execution flow, visualization minimized manual debugging methods, e.g. breakpoint tracing and console inspection. The integrated AI assistant also enhanced the efficiency of debugging by offering contextual explanations, and automatic syntax correction, which greatly minimized the amount of debugging effort required when using the system by beginner users. Visualization updates of the standard algorithms took 1-3 seconds, whereas recursive execution took about 6-8 seconds according to the size of the dataset and network conditions. Stability and compatibility were tested by using the Chrome, Edge, and Firefox browsers. The sandbox environment that was created using Docker ensuring that the execution is secure and did not influence responsiveness. Altogether, visualization, execution tracing, and AI-supported debugging increased the usability, efficiency of learning and understanding the program-flow in a browser-based programming environment.

4. DISCUSSION

The results of the experiments show that real-time visualization combined with AI-assisted debugging contributes to enhance the understanding of a program and lower the debugging effort of novice programmers. Docker-based sandboxing ensures controlled execution, and

trace streaming allows you to view the trace in the browser of your choice. The system, however, does only have support for a few programming languages and the use of third-party AI APIs can add latency. More languages, predictive code assistance, collaboration, integration with LMS, and version-control support are all areas that will be explored in the future.

5. CONCLUSION

NeuroFlow IDE brings AI-powered debugging, live code visualisation and safe container-based execution to one browser-based IDE. The concepts were better understood and less debugging was required when novice and intermediate programmers were evaluated experimentally. More language support, predictive help, collaboration and deployment on the cloud are planned for future work.

6. REFERENCES

[1] T. Phung, V.-A. Pădurean, A. Singh, C. Brooks, J. Cambronero, S. Gulwani, A. Singla, and G. Soares, "Automating human tutor-style programming feedback: Leveraging GPT-4 tutor model for hint generation and GPT-3.5 student model for hint validation", in Proc. 14th Int. Conf. Learning Analytics & Knowledge (LAK 24), Kyoto, Japan, Mar. 18–22, 2024, pp. 12–23. New York, NY: ACM.

[2] Z. Zhou, S. Wang, and Y. Qian, "Learning From Errors: Exploring the Effectiveness of Enhanced Error Messages in Learning to Program", Frontiers in Psychology, vol. 12, p. 768962, Nov. 30, 2021.

[3] K. K. K, V. Kondabathini, and Dr. B. Bharathi, "Sorting Algorithm Visualizer", Sathyabama Institute of Science and Technology, Chennai, India, Mar. 2021.

[4] M. Mukasheva, Z. Kalkabayeva, and N. Pussyrmanov, "Visualization of sorting algorithms in the virtual reality environment", Frontiers in Education, vol. 8, Article 1195200, Aug. 9, 2023.

[5] H. Aïdasso, M. Sayagh, and F. Bordeleau, "Build Optimization: A Systematic Literature Review", ACM Computing Surveys, vol. 58, no. 1, pp. 1–38, Jan. 2025.

[6] S. Shahaba, R. Razeen, A. Zubia, T. Surwadkar, V. Shah, and N. Shaikh, "Sorting Algorithm Visualizer", International Journal of Creative Research Thoughts (IJCRT), vol. 12, no. 2, pp. 2568–2577, Feb. 2024.

[7] J. Gong, V. Voskanyan, P. Brookes, F. Wu, W. Jie, J. Xu, R. Giavrimis, M. Basios, L. Kanthan, and Z. Wang, "Language Models for Code Optimization: Survey, Challenges and Future Directions", arXiv preprint arXiv:2501.01277, Jan. 2, 2025.

[8] T. Kräuter, H. König, A. Rutle, and Y. Lamo, "The Visual Debugger Tool," arXiv preprint arXiv:2404.12932, Apr. 19, 2024.

Biographies



Aniket Atram is presently pursuing his Bachelor's degree in Computer Engineering at Vishwakarma Institute of Technology, Pune. His interests include software development and emerging technologies.



Ansh Bagul is presently pursuing his Bachelor's degree in Computer Engineering at Vishwakarma Institute of Technology, Pune. His areas of interest include computer engineering and emerging technologies.



Yash Bhongale is presently pursuing his Bachelor's degree in Computer Engineering at Vishwakarma Institute of Technology, Pune. His interests include web development and programming.



Ruchika Birajdar is presently pursuing his Bachelor's degree in Computer Engineering at Vishwakarma Institute of Technology, Pune. She is interested in AI and software development.



Mrs. Shobha Mahendrakumar Patil is an Assistant Professor at the Vishwakarma Institute of Technology (VIT), Pune in the Department of Computer Engineering. She is a PhD holder from Bharati Vidyapeeth University (2025), a ME holder from Savitribai Phule Pune University (2015) and BE holder from Shivaji University (2008).

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Energy-Aware Network Coding AOMDV for IoT-Enabled MANETs

Vishal Bogam[1], Aarti Pimpalkar[2]

[1]Department of CSE, MIT ADT University, Pune, India, [2]Department of CSE, MIT ADT University, Pune, India. vishal.bogam@gmail.com, aartipimpalkar@gmail.com

Abstract.

With the number of increasing IoT devices within MANETs, the problems of energy efficiency, route stability, and reliable data delivery arise. Traditional multipath protocols are energy efficient, but introduce redundant paths and often require frequent route discovery, which adds to the energy consumption, and network-coding-based protocols may not be energy-aware. This paper introduces an Energy-Aware Network Coding AOMDV (EA-NC-AOMDV) protocol which incorporates the residual-energy-based path selection and adaptive network coding.

Keywords - Energy-Efficient Routing, Network Coding, AOMDV, IoT-Enabled MANET, Residual Energy Optimization.

1. INTRODUCTION

1.1. Background

The growing use of IoT in smart applications has increased the adoption of MANETs. However, IoT devices have limited energy, processing capacity, and transmission range [1–4]. AOMDV improves reliability through multipath routing but increases redundancy and control overhead, affecting energy efficiency [5–7]. Network coding improves throughput and reduces retransmissions by combining packets at intermediate nodes [8–9].

1.2. Research Gap

Existing approaches focus either on energy-aware routing [6], [12] or network coding [8], [10] separately. Very few studies tried to integrate both mechanisms in effective manner [13]. Also, traditional network coding schemes do not consider residual energy while taking routing decisions, which results in uneven energy consumption and premature node failures.

Key challenges include:

- High energy consumption due to frequent route maintenance [3], [4].
- High control overhead in multipath routing [5].
- Packet redundancy across multiple paths [6].
- Lack of energy awareness in network coding schemes [10].

1.3. Contribution:

1.4. Related Work:

Energy-aware MANET routing improves network lifetime by considering residual energy during route selection [2–4]. IoT-based MANET protocols reduce control overhead and battery consumption for resource-constrained devices [8]. However, existing approaches mainly focus on either energy optimization or network coding. Table 1.4.1 compares the proposed protocol with existing methods.

Table 1.4.1: Protocols comparison with parameters

Protocol	Energy (J)	PDR (%)	Delay (ms)	Lifeti me (s)
AOMDV	1280	87.2	142	160
NC-AOMDV	1185	90.5	135	170
Energy-AODV	1102	88.9	138	182
EA-NC-AOMDV	960	94.1	128	210

2. PROPOSED METHODOLOGY:

2.1. Network Model: The network will be composed of N IoT-enabled mobile nodes that will be randomly distributed within a 1000 × 1000 m^2 field. Nodes are based on the Random Waypoint mobility model [11] whose transmission range is uniform R and has a limited battery capacity. The intermediate nodes are allowed to conduct the coding operations based on predetermined conditions. Figure 2.1.1 shows this system flow.



Figure 2.1.1 System Flow of proposed methodology

2.2. Energy Model

The first-order radio energy model is implemented [1].

Transmission energy:

$$E_{tx}(k,d) = kE_{elec} + k\epsilon_{amp}\,d^2 \qquad (2.1)$$

Reception energy:

$$E_{rx}(k) = kE_{elec} \qquad (2.2)$$

Where:

k = packet size (bits), d = transmission distance, E_{elec}= electronic energy, ϵ_{amp} = amplifier energy

The model allows the proper calculation of the energy usage of nodes using equation (1) and (2).

2.3. The proposed EA-NC-AOMDV protocol:

The proposed protocol is called EA-NC-AOMDV, and it is recommended to decrease the cases of severe acute respiratory syndrome (SARS) outbreaks.

Step 1: Route Discovery:

The Route Request (RREQ) packets are changed to carry a field of residual energy. The energy value of each intermediate node is added. A composite cost function is:

$$\text{Cost} = \alpha \frac{1}{Residual\ Energy} + \beta\ \text{HopCount} \qquad (2.3)$$

Where α and β are weighting factors. Figure 2.3.1 shows the Route discovery in EA-NC-AOMDV.

Step 2: Energy based path selection.

The destination chooses disjointed paths of minimum composite cost value in k node to the node.



Figure 2..3.1. Route discovery in EA-NC-AOMDV

1. Step 3: Adaptive Network Coding:
2. Intermediate nodes perform XOR-based packet coding [7] only when the buffer threshold and minimum residual-energy conditions are satisfied, reducing unnecessary energy consumption [10].
3.
4. Algorithm 1: EA-NC-AOMDV
5.
6. Send RREQ packets with residual energy information.
7. Update cumulative path cost.
8. Select (k) optimal disjoint paths.
9. Enable coding when energy and buffer conditions are met.
10. Forward coded packets.
11. Update residual energy after each transmission.

3. EXPERIMENTAL RESULTS

3.1. Simulation Setup

Simulations are performed using NS-3 simulator considering IoT - oriented conditions [5].

Table 3.1. Experimental parameters with values

Parameter	**Value**
Area	1000 × 1000 m²
Nodes	50–100
Mobility	Random Waypoint
Traffic	CBR
Packet Size	512 bytes
Simulation Time	200 s

Compared protocols: AOMDV, NC-AOMDV, Energy-AODV and Proposed EA-NC-AOMDV

Energy Consumption: The total energy used by the nodes is the energy consumption [5]. Equation (4) shows the how to calculate PDR.

Packet Delivery Ratio (PDR):

$$PDR = \frac{Packets\ Received}{Packets\ sent} \qquad (3.1)$$

End-to-End Delay: Mean time to deliver packets.
Throughput: Bits per second delivered successfully.
Network Lifetime: Time of battery depletion of the first node [6].

3.2. Discussion:

Figures 3.2.1–3.2.5 compare the proposed EA-NC-AOMDV protocol with existing protocols. The proposed method achieves higher throughput, improves PDR by 10–12%, and reduces energy consumption by 18–22%, while also improving end-to-end delay and routing overhead.

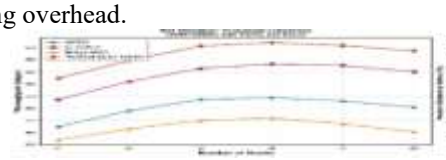


Figure 3.2.1. Throughput comarison

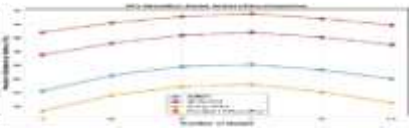


Figure 3.2.2. Packet delivery ratio comparison.

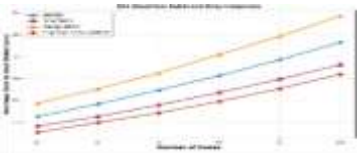


Figure 3.2.3. End to end delay comparison

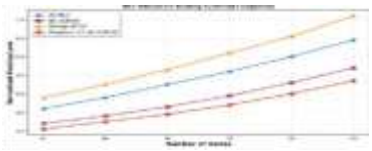


Figure 3.2.4. Routing overhead comparison

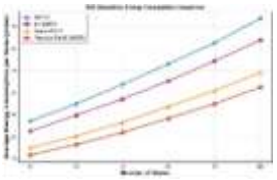


Figure 3.2.5. Energy consumption comparison

4. CONCLUSION

In this paper, EA-NC-AOMDV, a novel energy-aware multipath routing protocol that incorporates adaptive network coding to IoT-enabled MANETs is introduced. The protocol reduces up to 20 per cent of the energy used by combining residual energy and hop count with route selection, and permits conditional coding, resulting in dramatically reduced energy used in the network, as well as a much better ratio of packets delivered and an extended network lifetime. The throughput is also improved as compared to existing routing algorithms.

5. REFERENCES

[1] R. S. Sharma, "A review on energy-efficient routing protocols in wireless sensor networks," IEEE Access, vol. 8, pp. 163158–163181, 2020.

[2] T. Rana and D. Kumar, "IoT-enabled mobile ad hoc networks: Challenges and future trends," IEEE Access, vol. 10, pp. 34290–34307, 2022.

[3] H. Wang, L. Zhang, and K. Chen, "Energy-aware AODV routing protocol for IoT applications," IEEE Internet Things J., vol. 8, no. 4, pp. 2719–2731, Feb. 2021.

[4] Y. Yang, J. Cao, and X. Feng, "Residual energy based routing for wireless sensor and . IoT networks," IEEE Access, vol. 10, pp. 20456–20469, 2022.

[5] A. Sultana and M. A. Khan, "Residual energy based multipath routing for mobile adhoc networks," IEEE Access, vol. 9, pp. 123456–123467, 2021.

[6] L. Liu, X. Liu, and J. Zhang, "Energy-efficient routing with network coding in wireless ad hoc networks," IEEE Trans. Wireless Commun., vol. 22, no. 1, pp. 1020–1032, Jan. 2023.

[7] M. Z. Ali, M. A. Khan, and N. Javaid, "An efficient XOR network coding scheme for MANETs," IEEE Access, vol. 9, pp. 56789–56799, 2021.

[8] J. Li, X. Zhou, and B. Zhang, "Network coding-enhanced AOMDV for reliable MANET communications," IEEE Trans. Veh. Technol., vol. 70, no. 10, pp. 10589 10598, Oct. 2021.

[9] S. Kumar and M. Singh, “A survey of network coding in wireless ad hoc and sensor networks,” IEEE Commun. Surveys Tuts., vol. 23, no. 1, pp. 33–59, First Quart, 2021.
[10] Q. Zhao, J. Ma, and Y. Wang, “Adaptive network coding for energy balance in MANETs,” IEEE Trans. Mobile Comput., vol. 24, no. 5, pp. 4011–4023, May 2025.
[11] Saud Althunibat, Osamah S. Badarneh,Raed Mesleh,”Random Waypoint Mobility Model in Space Modulation Systems” IEEE Communications Letters Volume: 23 Issue:5,March2019
[12] K. Khan and W. Goodridge, “Energy Aware Ad Hoc On-Demand Multipath Distance Vector Routing,” Int. J. Intell. Syst. Appl., vol. 7, no. 7, pp. 50–56, 2015.
[13] M. Elhoseny, K. Shankar, and A. Singh, “Energy efficient routing in IoT-based MANETs using hybrid optimization techniques,” Wireless Networks, vol. 30, pp. 1457–1472, 2024.
[14] S. Sharma, A. Kumar, and R. Gupta, “A survey on network coding techniques for energy-efficient communication in MANETs,” IEEE Access, vol. 12, pp. 55678 55695, 2024.

Biographies



Vishal Bogam received bachelor’s degree in Computer Engineering in 2006 from University of Pune and master’s degree in Computer Engineering in 2017 from University of Mumbai. He is pursuing Ph.D. from MIT Art design and Technology University, Pune India. He has 15 years of teaching experience and 3 years of industrial experience. His area of interest is Networking, IOT, AI and ML.



Aarti P. Pimpalkar received her M.E. degree in Computer Engineering from Savitribai Phule Pune University, Pune, India, in 2013. She is currently pursuing her Ph.D. at MIT Art, Design and Technology University, Pune, India, and is working as an Assistant Professor at the same university. She has 15 years of teaching experience and 6 years of industry experience. Her research interests include AI and ML, Deep Learning, Image Processing, Visual Cryptography, and Cloud Computing.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Ask Your Trash: A Smart Bin System for Waste Segregation and Tracking

Toshita Bhagwate, Yoshita Bardhe, Aishwarya Mutha, Dr. Archana G. Said

Computer Engineering, AISSMS Institute of Information Technology, Pune, India,
toshitabhagwate@gmail.com, bardheyoshita@gmail.com,
muthaaishwarya19@gmail.com, archana.said@aissmsioit.org

Abstract.
This paper presents Ask Your Trash, an IoT-based waste management system with a database tracking layer that enables accountability via QR-based user registration. Unlike current smart bins, which mostly track waste fill levels, Ask Your Trash ensures complete auditability through linking of USER_ID with BIN_ID based on waste type (determined through classification by a CNN with 94.64% accuracy). This ensures creation of a continuous database entry. Hence, disposal of waste becomes a traceable process since every waste user is assigned an identity. Using MobileNetV2 (94.83% validation, 94.64% test accuracy on 3,080 unseen images), the system achieves 42 ms per image classification speed on a Raspberry Pi 4B. Comparatively, Custom CNN (77.08%), EfficientNetB0 (92.11%), and ResNet50 (92.09%) models do not exhibit comparable accuracy-speed tradeoff. With actual validation based on 500 images captured via the Pi Camera v2 in Pune, the model achieves 94.8% accuracy. The combined use of sensors (ultrasonic, load cell, moisture), CNN-based classification, QR-based identification and tracking, database storage, and mobile app-based visualizations results in up to 80% reduction in manual work and user accountability for waste sorting.

Keywords: Internet of Things; Smart Waste Management; Waste Segregation; QR Code Tracking; Smart City Infrastructure

1. INTRODUCTION

Urbanization and population expansion have resulted in an increased number of wastes that highlight the inefficiency of current waste management models. Traditionally, these methods rely on manual checks and scheduled collections, which do not suit dynamic urban environments [1], [2]. Hence, waste bins either overflow or are picked up while being still empty, creating inefficiencies and causing pollution problems [2]. Another substantial limitation is that information about waste bin capacity and composition is often unavailable and unobtainable. There are no insights into waste amounts and types, which results in inefficient service delivery. This problem makes it more difficult to segregate waste, which negatively impacts sustainability goals [3]. Development in IoT technologies has led to the creation of innovative smart waste bins that monitor bin capacity and send this data to the centralized system [1], [4]. However, at this stage, current approaches pay too much attention to technology rather than its users.The proposed “Ask Your Trash Smart Bin System” can resolve these problems by integrating sensor-based monitoring, QR code-driven user engagement, and centralized data management. This approach enhances transparency, optimizes collection, and promotes responsible waste disposal behaviour.

2. RELATED WORK

Waste management research has increasingly adopted digital technologies, beginning with IoT-based monitoring [1]. Early studies used sensors such as ultrasonic sensors to monitor bin fill levels and optimize collection schedules [2], [3]. Later systems introduced digital platforms to encourage user participation through reward-based waste disposal [3]. Wireless communication further enabled automated transmission of bin data to centralized dashboards, reducing manual inspections and supporting city-wide scalability [4], [5]. QR-code technologies improved secure and traceable user interactions [9] by linking physical waste disposal with digital user records [6], [7], [8], [10]. Building on these developments, the Ask Your Trash Smart Bin System integrates waste monitoring with individual user-level tracking. It records waste type and mass in a centralized database linked to user accounts, enabling users to view their disposal history through a mobile application. A planned reward mechanism will assign points based on waste categories and quantities, which can be redeemed through partner companies.

System	IoT Monitoring	User Interaction	AI Waste Classification
[2] IoT waste monitoring architecture	Yes	No	No
[3] Reward-based garbage monitoring	Yes	Yes	No
[8] IoT garbage monitoring framework	Yes	No	No
[10] Smart garbage monitoring using IoT	Yes	No	No
This work	Yes	Yes	Yes

Table 1. Analysis of the proposed architecture with existing research.

3. SYSTEM ARCHITECTURE

The proposed Ask Your Trash Smart Bin System uses a multi-layer architecture integrating IoT, AI, and software for automated waste classification, monitoring, and user tracking. The Smart Bin Layer includes ultrasonic, load cell, and moisture sensors, a camera, Raspberry Pi, and QR-based bin identification. The Intelligence Layer uses MobileNetV2 for image-based waste classification. The Application Layer provides a mobile application and admin dashboard for tracking disposal activities and monitoring waste data. The Data Management Layer stores and synchronizes user, bin, and waste information.

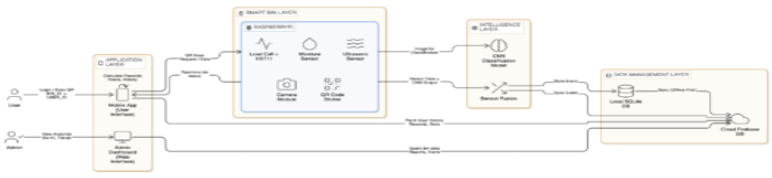


Figure 3.1. System architecture diagram

4. METHODOLOGY

The model was trained using public Kaggle datasets, including TrashNet, Garbage Classification Dataset, and Waste Classification Dataset, along with a custom dataset, comprising 20,506 images across five classes: cardboard, glass, metal, paper, and plastic.

The data were divided into 70% training, 15% validation, and 15% testing sets, with 3,080 test images. Images were resized to 224 × 224 pixels and normalized, while training augmentation included rotation, shifting, zooming, flipping, brightness adjustment, and shearing. MobileNetV2 was selected for its lightweight architecture and suitability for Raspberry Pi 4B deployment. Training used a two-phase Adam optimization strategy with categorical cross-entropy, initially freezing the base model and subsequently fine-tuning deeper layers. Data augmentation and class balancing improved generalization and minority-class performance.

5. MODEL EVALUATION

The model was evaluated using 5-fold stratified cross-validation on the full dataset and a holdout test set of 3,080 images (15%). Performance was assessed using macro-averaged Accuracy, Precision, Recall, and F1-score. Custom CNN, EfficientNetB0, and ResNet50 were trained under identical conditions for baseline comparison.

Model	Validation Acc.	Test Acc.	Pi 4B Inference	Parameter(M)
Custom CNN	81.17%	77.08%	90ms	5.1
Efficient-NetB0	93.93%	92.11%	67ms	5.3
ResNet50	94.21%	92.09%	112ms	23.5
MobileNetV2	94.83%	94.64%	42ms	2.3

Table 5.1.1. Performance comparison of baseline models and MobileNetV2

MobileNetV2 provides the best trade-off for edge deployment. Although EfficientNetB0 and ResNet50 achieve over 95% accuracy, their inference speeds are unsuitable for real-time Raspberry Pi 4B deployment. MobileNetV2 achieves 94.83% accuracy with a 42 ms inference time (23.81 FPS), while using 78% fewer parameters and providing 5.7× faster inference.

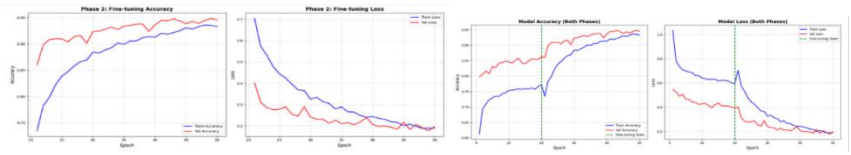


Figure 5.2.1 Model training performance

```
Per-Class Performance:
========================================
Class           Samples    Accuracy
----------------------------------------
cardboard       465          96.77%
glass           753          95.62%
metal           563          93.96%
paper           651          94.16%
plastic         648          93.06%
========================================
```

```
Detailed Classification Report:
============================================================
              precision    recall  f1-score   support

   cardboard      0.976     0.968     0.972       465
       glass      0.942     0.956     0.949       753
       metal      0.969     0.940     0.954       563
       paper      0.952     0.942     0.947       651
     plastic      0.907     0.931     0.919       648

    accuracy                          0.946      3080
   macro avg      0.949     0.947     0.948      3080
weighted avg      0.947     0.946     0.947      3080
```



Figure 4. Sample predictions on real-world test images, Mobile application interface, and Database

The SmartBin mobile application supports user authentication, QR-code bin scanning, waste tracking, reward management, and account administration. It uses Firebase Realtime Database for real-time monitoring of users, bins, and waste data. An offline-first architecture stores sensor data locally in SQLite and automatically synchronizes unsynced records with Firebase when connectivity is restored, ensuring data consistency and reliable monitoring.

6. RESULTS AND CONCLUSION

MobileNetV2 achieved the best performance with 94.64% accuracy and 24 FPS on Raspberry Pi 4B, outperforming EfficientNetB0 and ResNet50. Although plastic–glass misclassification remains a limitation, additional multi-angle training data can improve performance. The proposed Edge AI system enables automated waste sorting, reducing labor costs and improving recycling efficiency.

7. REFERENCES

[1] Eka Priyanto et al., "Smart Garbage Monitoring System Based on IoT and Cloud Computing," IEEE Sensors Journal, vol. 20, no. 11, pp. 5920-5930, 2020.

[2] K. N. F. K. Mustafa, “Smart BIN: IoT garbage monitoring system,” MATEC Web of Conferences, 2017.

[3] D. Aggarwal, S. Mittal, K. Upreti, and P. Nayak, “Reward based garbage monitoring and collection system using sensors,” Journal of Mobile Multimedia, pp. 391–410, 2024.

[4] B. Muthukumar et al., “QR code and biometric based authentication system for trains,” IOP Conference Series, 2019.

[5] J. Kim et al., “Implementation of QR code recognition technology using smartphone camera for indoor positioning,” Energies, vol. 14, 2021.

[6] R. Khan et al., “Smart garbage monitoring system using IoT,” SSRN, 2021.

[7] Anitha, “Garbage monitoring system using IoT,” IOP Conference Series, 2017.

[8] S. M. Chaware et al., “Smart garbage monitoring system using IoT,” IJIREEICE, 2017.

[9] Mark Sandler Andrew Howard Menglong Zhu Andrey Zhmoginov Liang-Chieh Chen Google Inc., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520.

[10] P. K. Owusu, “Smart garbage monitoring system using Internet of Things,” Middle East Journal of Applied Science & Technology, 2020.

Authors Biography



Toshita Bhagwate is a Computer Engineering graduate from AISSMS Institute of Information Technology, Pune, India. Her interests include Artificial Intelligence, IoT, and smart sustainable solutions.



Yoshita Bardhe is a Computer Engineering graduate from AISSMS Institute of Information Technology, Pune, India. Her interests include Artificial Intelligence, IoT, and emerging intelligent technologies.



Aishwarya Mutha is a Computer Engineering graduate from AISSMS Institute of Information Technology, Pune, India. Her interests include Artificial Intelligence, IoT, and practical smart solutions.



Dr. Archana G. Said is a faculty member in the Department of Computer Engineering at AISSMS Institute of Information Technology, Pune, India. Her academic interests include artificial intelligence, machine learning, IoT, and related emerging technologies.

A Design and Implementation of Jet shrouded Wind Turbine System

Ms. Deepali V. Wagh[1], Sumit V. Chim[2], Vrushali B. Mundphan[3], Vaishnavi V. Patil[4]

Department of Electrical Engineering, @ Bhivarabai Sawant Institute of Technology and Research, Wagholi, Pune, India,
arati.aher12@gmail.com; sumitchim1@gmail.com; mundphanv@gmail.com; vaishnavipatil2822@gmail.com

Abstract.

Scalable and sustainable energy sources are required for the entire world. When wind energy conditions are not predictable, the conventional wind turbine frequently faces difficulties. In order to investigate pressure distribution, velocity, and efficiency, this paper uses LABVIEW simulations to analyze the design and performance of jet turbine systems. The study examines how the system performance configuration greatly improves turbine efficiency when compared to traditional designs by modelling the shroud design of a jet turbine. This makes it a promising method for small, decentralized wind energy applications. The hardware-in-the-loop testing of the jet wind turbine system is based on this work. LabVIEW simulations were used for performance analysis to assess torque generation, power output, velocity enhancement, and pressure distribution.
Document of your own.

Keywords: Jet Wind Turbine; Wind Energy; LabVIEW Simulation; Renewable Energy; Energy Efficiency.

1. INTRODUCTION

The growing demand for clean energy has increased interest in wind power. Conventional wind turbines have limited efficiency under low and variable wind speeds due to turbulence, yaw misalignment, and insufficient airflow. This study proposes a compact jet wind turbine that accelerates airflow to improve energy extraction while requiring less installation space. Since wind power increases with the cube of wind speed, enhancing airflow velocity can significantly improve turbine output.

2. LITERATURE SURVEY

Previous studies have explored shrouded and vertical-axis wind turbines to improve wind energy extraction [1–2]. Research has also examined yaw control, turbulence effects, blade shapes, and generator parameters for improved efficiency [3–5]. 3D modelling and printing techniques have been used for optimized turbine designs [6], while helical-blade turbines have shown potential for efficient electricity generation in rural and emergency applications [7].

3. JET WIND TURBINE SYSTEM

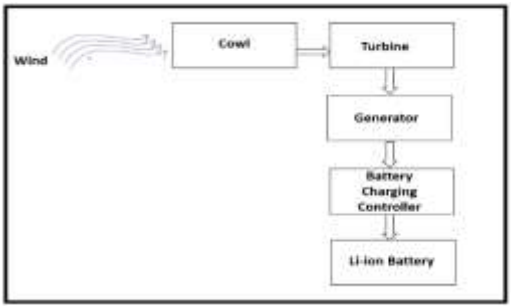


Figure 3.1. Block diagram of Jet Wind Turbine System

In this proposed system wind is blow with 20m/s and directed towards the cowl shape wind turbine. The cowl shape structure utilizes the more wind energy which is input to the jet

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

wind turbine. The generator generates the power which is controlled by TP4056 battery charging controller. The output is tested by connecting the two series Li-Ion rechargeable batteries with successfully charged up to 5.6V.

Work	Flow Acceleration	Compact Design	Simulation Validation
Variable Speed [1]	NO	NO	YES
Wind Lens Turbine [2]	YES	NO	YES
Blade Optimization [5]	NO	NO	YES
Cowl shape jet turbine	YES	YES	YES

Table 3.1.1. Comparative Analysis of Jet wind Turbine

3.1. Calculations

The system consists of,

i. Converging duct (cowl)
ii. Cambered rotor blades
iii. Shaft coupling
iv. Generator
v. Li-ion battery
vi. Battery Charge Controller

Power equation used:

$$P = \frac{1}{2}\,\rho A v^3 Cp$$

Where,

ρ= Air Density
A = Swept Area
v = Wind Velocity
Cp = Power Coefficient

3.2 Theoretical Calculation:

In conventional Wind Turbine,
v = 20 m/s
Diameter = 0.27 m
ρ = 1.2 kg/m³
Calculated Power ≈ 274 W

3.3 Efficiency improvement:

Jet turbine aerodynamic efficiency:0.45
Jet turbine mechanical/electrical efficiency:0.9
Conventional turbine aerodynamic efficiency:0.35
Conventional mechanical/electrical efficiency:0.8

Efficiency improvement = $\frac{0.45*0.9}{0.35*0.8} = 1.45$

For Conventional wind turbine Power

output = $274 * 0.28 = 76.7Watt$

For Jet Wind Turbine Power output = $274 * 0.405 = 111Watt$

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

4. RESULT AND DISCUSSION

4.1 Simulation Results:

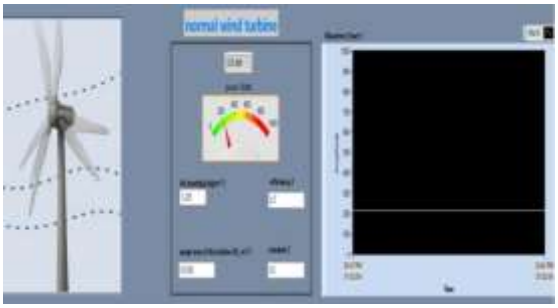


Figure 4.1. Conventional Wind Turbine Simulation in LABVIEW

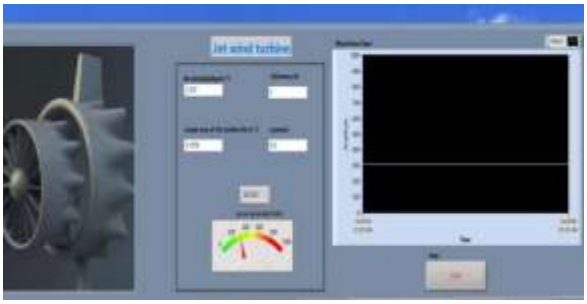


Figure 4.2. Jet wind turbine Simulation in LABVIEW

4.2 Hardware Results:

The shrouded configuration lowers turbulence losses and greatly increases airflow concentration. Figure 4 shows the prototype model of jet wind turbine.

Overall efficiency by about 1.45 times when compared to conventional turbines under the same environmental conditions. In this model, two 3.7 V rechargeable batteries are connected in series to show the working of the jet wind turbine and to store power in a stable way.



Figure 4.3. Prototype of Jet Wind Turbine



Figure 4.4. Discharge Battery



Figure 4.5. Charged Battery

The system shows higher torque, lower turbulence losses, and better stability. Rooftop installations, telecom towers, urban renewable energy systems, and hybrid solar-wind energy systems are examples of applications. Because wind speed and power have a relationship, the results verify that airflow acceleration significantly boosts power output. The jet configuration stabilizes output performance and enhances torque generation.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Deployment in rooftop and urban installations is made possible by the structure's compact size. Never the less, additional experimental validation and analysis are needed.

5. CONCLUSION

Conventional wind turbines are less effective in low-wind-speed regions. This study proposes a jet wind turbine that accelerates airflow to improve energy capture. LabVIEW-based monitoring showed approximately 1.45× higher efficiency than a conventional turbine, with successful battery charging, demonstrating its potential for low-wind applications.

6. REFERENCES

[1] K. Naoi & M. Shiono, 'Relationship between incident angle of wind on rotor blade and output of a drag-type multi-blade vertical-axis wind turbine', Proc. IEEE Conference on Power, Energy & Electrical Engineering,2021.

[2] K. Shimizu et al., 'Improvement of generating efficiency of vertical-axis wind turbine with wind lens', Proc. IEEE Global conference on Consumer Electronics,2022.

[3] Q. Zou & Z. Tang, 'Analysis and optimization of wind turbine yaw regulation strategy,' in Proc. IEEE IRCE, 2023

[4] M. V. H. G. dos Santos, T. C. da Silva & J. C. Teixeira, 'An estimation model for the turbulence index using wind speed averages in wind turbines', IEEE DUSCON ,2023.

[5] M. Debbache & F. Meziane, 'Design and sizing of wind blades based on generator parameters for small wind turbines for low wind regions', IEEE International Conference on Electronics, Energy and Measurement, 2023.

[6] H. Hu et. al., 'Experimental wind tunnel study of vertical axis wind turbine interactions', IEEE Advanced Electrical and Energy Systems, 2024

[7] H. S. Gadakh, R. P. Jawale, V. Kumar & D. B. Pardeshi ,'Development of helical type wind turbine', IEEE International conference on Inventive Systems and Control,2024

Biographies



Ms. Deepali V. Wagh received the bachelor's degree in Electrical Engineering from SPPU University in 2010, the master's degree in Electrical Power System engineering from SPPU University in 2020, respectively. She is currently working as an Assistant Professor at the Department of Electrical Engineering, Faculty of Engineering, JSPM's Bhivarabai Sawant Institute of Technology and Research, Pune. Her research areas include Smart Grid, Sustainable Energy, and Control System.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International



Mr.Sumit V.Chim received the bachelor's degree in Electrical Engineering from JSPM's Bhivarabai Sawant Institute of Technology and Research, SPPU University in 2026. He is currently working as GET in Archeral Pvt. Ltd. Mumbai. His research areas include Power Electronics, Renewable Energy, and Advanced Power System.



Ms. Vrushali B. Mundphan received the bachelor's degree in Electrical Engineering from JSPM's Bhivarabai Sawant Institute of Technology and Research, SPPU University in 2026. She is currently working as GET in ENGIE Pune. Her research areas include Deep learning, Sustainable Energy, and Smart Grid.



Ms. Vaishnavi V.Patil received the bachelor's degree in Electrical Engineering from JSPM's Bhivarabai Sawant Institute of Technology and Research, SPPU University in 2026. Her research areas include Cloud Computing, Renewable Energy, and Power system.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

<Conference abbreviation> <Volume number and Year> <DOI Number>
<Conference Series name>

Real-Time Tomato Leaf Disease Detection Using YOLOv7

Aditya Ramkrushna Mahale Puja Somnath Mahale Dr. A. S. Thorat Piyusha Rajendra Mahajan Adarsh Pravin Patil

Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner
adityamahale53@gmail.com,pujamahale733@gmail.com,abhinav.th1990@gmail.com
piyumaha48@gmail.com aadarshpatil29604@gmail.com

Abstract.

The devastating effects of tomato leaf diseases on crop yield and quality are so significant that it is necessary to detect early and accurately disease causes to provide good farm management. This study develops a real-time disease detection model from scratch using YOLOv7-tiny object detection. The model is trained with a dataset including 10 classes of 20,000 images from Roboflow Universe, including Bacterial Spot, Early Blight, Late Blight and Healthy leaves. The dataset has been divided in training, validation and testing set for good evaluation. As a last point, the model we propose has precision = 0.768, recall = 0.658 and mAP@0.5. Aside from detecting disease, the system estimates the severity of the disease and uses a web application to send an image along with automatic predictions. Findings showed the proposed method could be useful and relevant both for real time and applied use in agriculture, especially under conditions of scarcity of resources.

Keywords. Tomato leaf disease detection, YOLOv7-tiny, object detection, deep learning, precision agriculture, real-time detection.

1. INTRODUCTION

Tomato is a widely cultivated crop, but diseases such as Bacterial Spot, Early Blight, Late Blight, and Mosaic Virus can reduce yield and quality. Timely detection is essential for effective crop management. Conventional diagnosis relies on manual inspection and expert knowledge, limiting scalability. Computer vision and deep learning, particularly Convolutional Neural Networks (CNNs), have improved automated disease detection [1–4]. Other studies have explored feature-based methods, transfer learning, and advanced detection frameworks [5–7]. However, many approaches focus mainly on classification and provide limited disease localization [8,9]. YOLO models offer efficient real-time classification and localization [10].

Therefore, this study proposes a YOLOv7-tiny-based system for multi-disease detection and web-based prediction.

2. RELATED WORK

Machine learning and deep learning have been widely applied to plant disease detection. Mohanty et al. [4] demonstrated effective CNN-based plant disease classification on controlled datasets. Wang et al. [11] improved tomato disease detection using multi-scale feature fusion and attention mechanisms, while Pantazi et al. [6] explored image feature-based one-class classifiers. Phan et al. [7] applied YOLOv5 for real-time detection but provided limited localization. Karimanzira [2] emphasized context-aware detection for operational environments. However, lightweight models enabling real-time localization and classification on resource-constrained hardware remain limited.

Table 2.1: Comparison of Related Works with the Proposed Method

Method	Localization	Multi-Class	Real-Time	Deployment
[5]	Yes	Yes	Yes	Limited
[9]	No	Yes	No	No
[11]	Yes	Yes	Yes	Limited
[6]	No	Yes	No	No
This work	Yes	Yes	Yes	Yes

To address this limitation, this study proposes a lightweight YOLOv7-tiny model for multi-class tomato leaf disease detection, optimized for real-time deployment on resource-constrained hardware. Table 2.1 compares the proposed approach with existing methods in localization, multi-class support, real-time performance, and deployment suitability. The main contributions include: (1) disease classification and infected-region localization using bounding boxes; (2) disease severity estimation from detection outputs; and (3) integration into a Flask-based web application for real-time image upload and prediction.

3. METHOD, EXPERIMENTS AND RESULTS

3.1 Methodology

A YOLOv7-tiny object detection model was developed for tomato leaf disease detection using transfer learning with pre-trained weights to improve convergence and reduce training time. Data augmentation, including HSV adjustments, horizontal flipping, mosaic, and mixup, improved robustness to variations in lighting, occlusion, and leaf orientation. The model predicts disease classes with bounding boxes and class probabilities. Localization performance is evaluated using Intersection over Union (IoU), a standard object detection metric [10]. The complete system architecture is presented in Figure 4.1.

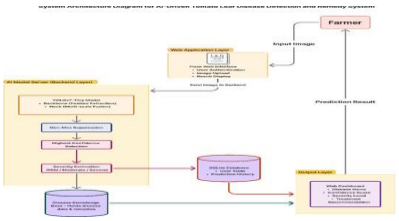


Figure 4.1. Complete system design for the YOLOv7-based tomato leaf disease detection system

The dataset contains approximately 20,000 annotated tomato leaf images across 10 disease classes in YOLO format. YOLOv7-tiny was trained for 25 epochs on an NVIDIA RTX 2050 GPU (4 GB VRAM) using batch size 2. SGD optimization used a learning rate of 0.001, momentum of 0.937, and weight decay of 0.0005. Performance was evaluated using precision, recall, mAP@0.5, and mAP@0.5:0.95.

3.2 Results

The YOLOv7-tiny model achieved a precision of 0.768, recall of 0.658, mAP@0.5 of 0.735, and mAP@0.5:0.95 of 0.465. Septoria and Late Blight showed higher AP values of 0.884 and 0.880, respectively, due to their distinct visual patterns. Leaf Miner achieved a lower AP of 0.456 because of its fine textures and smaller lesions. The confusion matrix and precision-recall curve are presented in Figures 4.2. Overall, the model provides reasonable multi-disease detection performance and inference efficiency on resource-limited hardware, although the 416 × 416 input resolution limits sensitivity to fine-grained features.

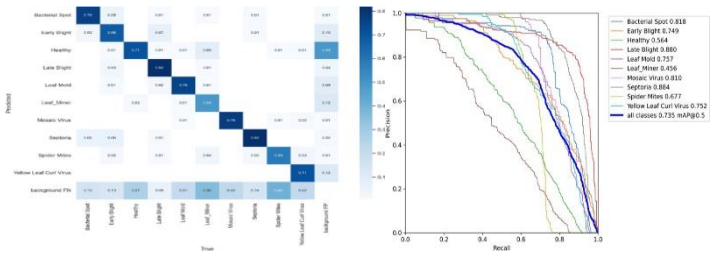


Figure 4.2. Confusion matrix and Precision–Recall curves.

4. CONCLUSIONS

The study addresses tomato leaf disease diagnosis, where manual inspection is time-consuming and dependent on expert knowledge. A transfer-learning-based YOLOv7-tiny model was developed for simultaneous disease classification and localization across ten disease classes. The model achieved 0.768 precision, 0.658 recall, and 0.735 mAP@0.5, with better performance for visually distinct diseases. Lower accuracy was observed for fine-textured diseases such as Leaf Miner. The 416 × 416 input resolution limits detection of small lesions and subtle patterns. Future work should explore higher-resolution inputs, advanced augmentation, improved architectures such as YOLOv8 or transformer-based models, field validation, and mobile deployment.

5. REFERENCES

[1] K. P. Ferentinos. Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145:311–318, 2018.

[2] D. Karimanzira. Context-aware tomato leaf disease detection using deep learning in an operational model. Electronics, 14(4):661, 2025.

[3] J. Liu and X. Wang. Tomato diseases and pests detection based on improved yolov3 convolutional neural network. Frontiers in Plant Science, 11:898, 2020.

[4] S. P. Mohanty, D. P. Hughes, and M. Salathe. Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7:1419, 2016.

[5] M. Umar. Precision agriculture through deep learning: Tomato plant multiple diseases recognition with cnn and improved yolov7. IEEE Access, 12:49167–49180, 2024.

[6] M. C. Pantazi, D. Moshou, and A. A. Tamouridou. Automated leaf disease detection in different crop species through image features analysis and one-class classifiers. Computers and Electronics in Agriculture, 156:96–104, 2019.

[7] Q.-H. Phan, T. N. Nguyen, and H. V. Le. Classification of tomato fruit using YOLOv5 and convolutional neural network models. Plants, 12(4):790, 2023.

[8] P. Singh, R. Kaur, and H. Sharma. Explainable ai in agriculture: A model for interpretable plant disease detection using grad-cam. IEEE Access, 10:125980–125992, 2022.

[9] L. Tan, J. Lu, and H. Jiang. Tomato leaf diseases classification based on leaf images: A comparison between classical machine learning and deep learning methods. AgriEngineering, 3(3):463–476, 2021.

[10] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022.

[11] Y. Wang, P. Zhang, and S. Tian. Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion. Frontiers in Plant Science, 15, 2024.

Biographies



Aditya Ramkrushna Mahale received the bachelor's degree in Computer Engineering from Amrutvahini College of Engineering, Sangamner, Maharashtra, India. His research areas include artificial intelligence, deep learning, computer vision, and real-time object detection, with a particular focus on developing intelligent solutions for agricultural applications.



Puja Somnath Mahale received the bachelor's degree in Computer Engineering

from Amrutvahini College of Engineering, Sangamner, Maharashtra, India. Her research interests include deep learning, computer vision, generative artificial intelligence, and real-time plant disease detection, with a particular focus on developing intelligent solutions for agricultural applications.



Dr. Abhinav Sudhir Thorat is an Assistant Professor in the Department of Computer Engineering at Amrutvahini College of Engineering, Sangamner, Maharashtra, India. He has over 12 years of teaching experience. His academic and research interests include Data Science, Big Data Analytics, Artificial Intelligence, Machine Learning, Generative AI, Data Mining, and Data Warehousing.



Piyusha Rajendra Mahajan received the bachelor's degree in Computer Engineering from Amrutvahini College of Engineering, Sangamner, Maharashtra, India. Her research interests include deep learning, computer vision, and plant disease detection.



Aadarsh Pravin Patil received the bachelor's degree in Computer Engineering from Amrutvahini College of Engineering, Sangamner, Maharashtra, India. His research interests include deep learning, computer vision, and real-time plant disease detection.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Integrated System for Leaf Disease Prediction and Crop Management

Subodhini Gurav

Department of Computer Engineering Dr. D. Y. Patil Institute of Technology Pimpri, Pune, India. subodhini01998@gmail.com

Abstract.

Plant diseases are an issue for people who grow plants everywhere. People who grow plants for a living have to deal with these plant diseases. It affects their lives. Plant diseases can be really bad, for people who need plants to survive. The ways we use now to find plant diseases are not very good because they need someone who knows a lot about plants to look at them. This makes it hard for farmers in away places to know if their plants are sick. This paper is about making a system that can look at a picture of a leaf and tell us what is wrong with the plant. We use something called Convolutional Neural Networks to do this. We do not need any equipment to help us we just need a picture of the leaf. This system can look at the picture. Tell us what disease the plant has. This is a good solution for farmers because it can help them know what is wrong with their plants and they can make their plants healthy again. Plant Disease Recognition is a deal and this system can help with that. The Plant Disease Recognition System is very good, at analyzing a leaf image and recognizing a disease.

Keywords. plant disease classification, convolutional neural network, deep learning, PlantVillage, precision agriculture, leaf image analysis

1. INTRODUCTION

Agriculture plays a critical role in economics worldwide. Among the challenges the sector faces plant diseases remain one of the most challenging threats to crop productivity and food security. Plant diseases significantly reduce crop yield and quality, directly impacting the livelihoods of farmers globally.

Traditionally, farmers inspect plant leaves visually to assess their health. However, this approach often fails to detect diseases at early stages, as leaves may appear healthy externally while already being infected. Today, advances in artificial intelligence provide powerful tools for automated image-based analysis. Deep learning convolutional neural networks (CNNs), a type of model, examine leaf images to detect and recognize them abnormal patterns with high accuracy [1].

2. LITERATURE REVIEW

2.1 Machine Learning based Agriculture Systems

With machine learning being one of the best ways of analyzing patterns and predicting results, it has been widely adopted in the agriculture sector. Machine learning has been shown to be an excellent resource when it comes to such aspects as yield forecasting, weather prediction, and nutrient testing. Nonetheless, machine learning has had the greatest success so far in diagnosing plant disease using image recognition. The convolutional neural network is primarily designed for working with visual information, whereby it recognizes patterns from the information provided. CNN models have been widely used for detecting plant diseases from leaf images with high accuracy.

2.2 Existing Approaches in Disease Prediction

The early approaches that made use of ML for diagnosing plant disease focused on classical machine learning techniques, including Support Vector Machines (SVM), Decision Tree, and Random Forest. Classical approaches needed manual extraction of features from the images to be used by the algorithm; in essence, human experts were responsible for defining such features as texture, colors, and shapes. However, this approach performed poorly when working with large datasets and was incapable of detecting small differences between diseases with visual similarities. The CNN algorithms have become preferable to classical ML algorithms due to their ability to automatically extract features. Traditional approaches also include image segmentation and soft computing techniques for disease detection. CNN-based models have also been successfully applied to classify potato leaf diseases.

2.3 Research Gap

Most of the existing methods include the use of IoT sensors, soil analysis, weather conditions, and crop suggestions, along with disease prediction. However, despite their usefulness, such methods can be costly, complicated, and difficult for farmers to adopt. There is a need for simpler methods that only use machine learning for predicting diseases based only on leaf images. The performance of deep learning models depends on dataset quality and environmental factors [2].

3. METHODOLOGY

In the proposed model, an image-based machine learning technique is adopted to identify diseases in plants. Only leaf images are considered, and there is no requirement to employ any IoT sensor devices. This makes the approach simpler, scalable, and easily accessible to farmers.

3.1 System Overview

This is accomplished by the use of a well-designed pipeline that resembles a natural leaf inspection process but with much more accuracy through the deployment of machine learning approaches. This is achieved by having the user upload an image taken on their mobile device or camera of a particular plant leaf. Upon upload into the system, a number

of pre-processing stages are performed to transform the image into a machine learning compatible format.

3.2 Dataset Description

The model is developed using the PlantVillage dataset, which contains over 54,000 images depicting both healthy and diseased plants from various crops. This dataset's diversity allows the model to identify unique features of different disease categories.

3.3 Data Preprocessing

Pre-processing of data is significant phase ensuring that the image is processed in order for the model to use them effectively. In order to maintain uniformity among all images, the dimensions of each image is standardized to 224×224 pixels. Normalization of pixel values further helps in enhancing the learning process of the model. Apart from this, there are other data augmentation techniques which involve operations like rotating, flipping, zooming, and adjusting brightness of images in order to improve the performance of the model.

3.4 CNN Architecture

The CNN model used in this paper employs several layers that learn increasingly complex features from the images. The initial convolution layers receive images as inputs and detect local features such as color gradients and edge patterns. These layers produce feature maps that represent various visual characteristics extracted from the input image.

3.5 Model Training

The model is trained using the Adam Optimizer, with a learning rate set at 0.001 and a batch size of 32, over the course of 40 epochs. The dataset is divided into training, validation, and testing sets in proportions of 70%, 15%, and 15%, respectively.

The backpropagation process minimizes cross-entropy loss function for multi-class classification.

3.6 Model Validation and Performance Metrics

After the model has been successfully trained, the next step will be evaluating its effectiveness. It will be performed using the metric of accuracy, which provides an overview of results obtained but in case some difficulties arise during the classification of specific classes, we will need metrics of precision, recall, and f1 score[3].

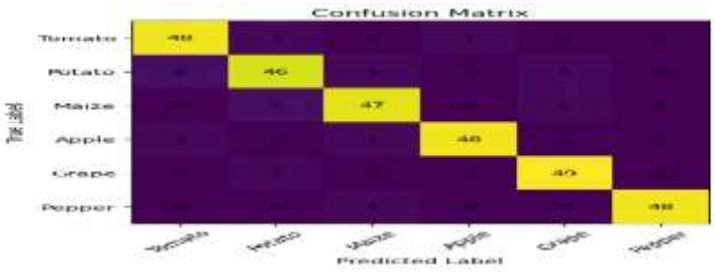


Figure. 2. Confusion Matrix of CNN Model

4. SYSTEM DESIGN

The proposed system's design is composed of four distinct modules that contribute to the processing of input data to generate the final diagnosis. The first module includes the process of inputting image data from users and the second one involves pre-processing the image before analyzing it. The essential component of the implemented system comprises the CNN prediction engine which performs the analysis of the image and generates predictions, and the third module involves converting these predictions into useful outputs [4].

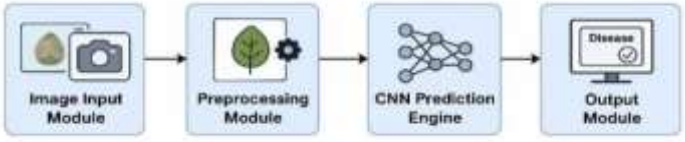


Figure.3. System Architecture of Proposed Model

5. RESULT AND ANALYSIS

As per the detailed study of the CNN, the following results are obtained; that is, it shows a level of accuracy of about 94.32%. This means that the network has a very high level of accuracy in the classification of various diseases. From the validation graph, it is clear that there is no case of overfitting, as there is a correlation between the accuracy levels [5].

Class	Precision	Recall	F1-Score
Tomato	0.95	0.94	0.94
Potato	0.93	0.92	0.92
Maize	0.94	0.93	0.93
Apple	0.95	0.94	0.94
Grape	0.96	0.95	0.95

Table 1. Performance Metrics of CNN Model

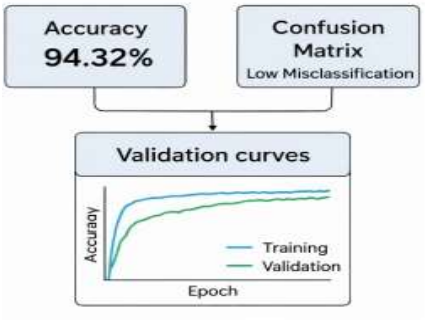


Figure. 4. Performance Evaluation Results

6. LIMITATIONS

Even though the proposed algorithm gives highly accurate results, it was mainly developed using the PlantVillage dataset consists of images captured in controlled environments. Real-life situations can be affected by many environmental conditions, like lighting changes, background noises, and leaf obstruction, among others. The next step for future works would be to train the model using real field images [6]

7. REFERENCES

[1] J. G. A. Barbedo, “Factors influencing the use of deep learning for plant disease recognition,” *Biosystems Engineering*, vol. 172, pp. 84–91, 2018.

[2] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, “Identification of rice diseases using deep convolutional neural networks,” *Neurocomputing*, vol. 267, pp. 378–384, 2017.

[3] S. Ramesh, R. Vydeki, and M. Senthilkumar, "Application of machine learning in diagnosing maize leaf disease using deep learning-based CNN architecture," Ecological Genetics and Genomics, vol. 21, p. 100089, 2021.

[4] H. Zeng, H. Yang, J. Zhang, L. Zhao, and D. Tu, "Identification of maize leaf diseases using improved deep convolutional neural networks," IEEE Access, vol. 9, pp. 159800–159812, 2022.

[5] N. Sharma, R. Jain, and A. Mishra, "An efficient deep learning-based plant disease detection system using transfer learning and data augmentation," Neural Computing and Applications, vol. 35, no. 8, pp. 5877–5893, 2023.

[6] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, "Detection of potato disease using image segmentation and deep learning for smart agriculture," IEEE Access, vol. 12, pp. 10234–10248, 2024.

[7] R. Verma, P. Gupta, and S. Tiwari, "Lightweight CNN-based mobile application for real-time plant disease diagnosis in precision agriculture," Computers and Electronics in Agriculture, vol. 220, p. 108845, 2025.

Biographies



Subodhini Gurav received the bachelor's degree in computer engineering from Shivaji University in 2022, the master's degree in computer engineering from Savitribai Phule Pune University in 2026, respectively.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Bridging the Semantic Gap: Natural Language to Relational Algebra and SQL Query Generation

Kirti H. Wanjale, Pragati A. Bandagar, Manasvi S. Bandkar, Aadi A. Bhangdiya, Yogini P. Bhamare

Vishwakarma Institute of Technology Pune, Maharashtra, India, kirti.wanjale@vit.edu, pragati.bandagar24@vit.edu, manasvi.bandkar24@vit.edu, yogini.bhamare24@vit.edu

Abstract.

There are Natural Language to SQL systems that have an immense contribution towards easing database access to users in the form of using their ordinary language for queries without needing a complicated syntax. On the other hand, conventional database management systems require the use of Structured Query Language which poses an obstacle to the use of these systems by non-programmers. In order to solve this problem, this paper presents an implementation of a system that will provide a connection between natural language, relational algebra, and SQL. The suggested method will follow a process-oriented approach in which the user's query is processed to obtain its relational algebra equivalent after which it will be translated to a suitable SQL command. Such a connection will help to make more clear the link between the intention of the user and the operation to be performed in the database. The experiment results show that there is improvement in the accuracy and clarity of the translation process from natural language to SQL.

Keywords. Natural Language, Structured Query Language (SQL), Relational Algebra (RA), Semantic Parsing, Query Generation

1. INTRODUCTION

The growing volume of data has increased the need for accessible database querying. Traditional SQL requires technical expertise, limiting database access for non-technical users. Natural Language-to-SQL (NL2SQL) systems address this challenge by enabling queries through natural language, although translation remains complex. Existing machine learning and semantic parsing approaches [2], [6] often lack transparency and interpretability. This study proposes a systematic mapping between Natural Language Queries (NLQ), Relational Algebra (RA), and SQL to improve traceability and understanding. The model also supports teaching NLP, RA, and SQL concepts [7].

2. RELATED WORK

Many attempts have been made to create NLTOSQL approaches to simplify the database access process. The early attempts were through rule-based approaches [6], which were

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

difficult to use when processing complex queries. Later attempts included neural and semantic parsing approaches [4], which offered more accurate results, albeit lacking explainability [1-10]. The current state-of-the-art NLTOSQL approaches include approaches like QuerySpeak [2] and the ones involving chatbots [3]. Techniques based on LLMs [1], [10] perform better, whereas some recent studies also use relational algebra [7], [8] for better explainability.

Table 1. Comparison of existing work with proposed work

Reference	NLP	RA	Transpar-ency
[1]	Yes	No	No
[2]	Yes	No	No
[3]	Yes	No	No
[4]	Yes	No	No
[7]	No	Yes	Yes
[8]	Yes	Yes	Yes
This work	Yes	Yes	Yes

3. METHOD, EXPERIMENTS AND RESULTS

The proposed system uses a Java-based multi-stage framework to convert natural language questions into SQL queries. First, the user query is normalized, and relevant entities and keywords are identified and semantically tagged. The system then checks the query for ambiguity and consistency. Valid queries are converted into relational algebra and corresponding SQL statements. The Application Layer uses an NLP parsing engine to interpret user intent and extract relevant information, which is passed to the Relational Algebra/SQL Generator for query construction. The generated SQL is executed through Java Database Connectivity (JDBC), which connects the application to the MySQL database. Retrieved results are processed and presented through the Presentation Layer, providing users with an accessible interface for submitting natural language questions and receiving database results.

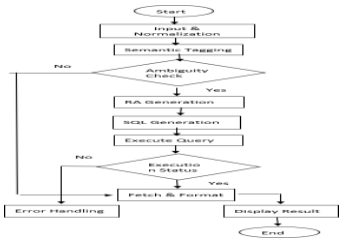


Figure 4.1. System Flowchart

4.1 Functional Illustration

The Design system produces many complete NLQ to SQL query conversions through relational algebra. Some of those examples are included in Table 2 below.

Table 2. Sample Outputs

Component	Example Output
NLQ	(1) show doctors specialization cardiology (2) show recent 10 appointments
SQL Query	(1) SELECT * FROM doctor WHERE LOWER(specialization) LIKE 'cardiology' (2) SELECT * FROM appointment ORDER BY visit_day DESC LIMIT 10
RA Expression	(1) σ condition (π * (doctor)) (2) τ (appointment)

The system improves traceability and transparency by generating Relational Algebra (RA) and SQL simultaneously, allowing users to follow the translation from logic to executable code [14][19]. MySQL compatibility through JDBC supports reliable query execution, while NLP processing handles linguistic variations effectively [12]. Its modular Java architecture further supports reliability and easier debugging [14]. The system also serves as a teaching instrument by demonstrating the relationship between NLQ, RA, and SQL [19][14]. Its modular design enables future extensions, including semantic relationships and query optimization [1]. This combination improves usability, learning, maintainability, and functional extensibility.

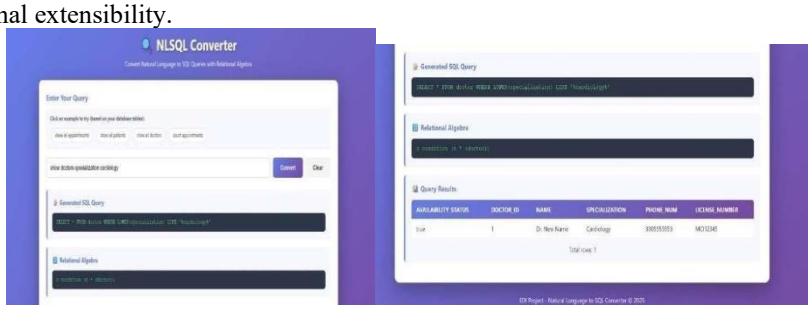


Figure 4.2. Results of the system

The developed NL2SQL system effectively bridges natural language queries and database processing through relational algebra and SQL. Relational algebra improves transparency by allowing users to trace query processing. The system performs reliably for straightforward to moderately complex queries, although complex or ambiguous queries require improved semantic analysis. Overall, the system supports efficient database access while providing educational value for understanding query processing.

4. CONCLUSIONS

This work addresses database access challenges for non-technical users by translating natural language into relational algebra and SQL queries executable through MySQL using JDBC. The system generates precise, understandable, and traceable queries while retrieving

relevant database results. However, it currently supports simple queries. Future work will focus on complex queries, improved semantic accuracy, and query optimization.

5. REFERENCES

1. A. K. Sharma, S. C. Kanumuri, P. AR, and S. S. Sharath, "Advancing Natural Language to SQL: A Comparative Study of Open Source LLMs on Benchmark Datasets," *2025 IEEE Symposium on Computational Intelligence in Natural Language Processing and Social Media (CI-NLPSoMe Companion)*, 2025.
2. T. S. V. Reddy, T. Charithat, T. Sathwik, S. V. Samitha, and N. Panda, "QuerySpeak: From Human Language to Database Queries," *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)*, 2024.
3. M. S. Mahagave et al., "AI Enabled Chat-Driven Database Management for NLP to SQL Conversion,"
   2025 Seventh International Conference on Computational Intelligence and Communication
4. R. Xu and A. Singh, "Neural Semantic Parsing Natural Language into SQL," *arXiv:2106.13858v1 [cs.CL]*, 2017.
5. D. M. L. Nethravathi et al., "Structuring Natural Language to Query Language: A Review,"
   Engineering, Technology & Applied Science Research, 2020.
6. G. Singh and A. Solanki, "An algorithm to transform natural language into SQL queries for relational databases," *Selforganizology*, 2016.
7. R. Suryawanshi et al., "Mapping of Relational Algebra Into SQL Queries," *Project Report, Vishwakarma Institute of Technology*, 2025.
8. F. S. C. Tseng, A. L. P. Chen, and W. P. Yang, "On mapping natural language constructs into relational algebra through E-R representation," *Information Sciences*, 1993.
9. A. Giordani and A. Moschitti, "Corpora for Automatically Learning to Map Natural Language Questions into SQL Queries," *ACL Anthology*, 2010.
10. X. Liu et al., "A Survey of Text-to-SQL in the Era of LLMs: Where are we, and where are we going?,"
    arXiv:2408.05109v5 [cs.DB], 2025.

Biographies



Dr. Kirti H. Wanjale is a Professor in the Department of Computer Engineering at Vishwakarma Institute of Technology, Pune. Her research interests include Artificial Intelligence, IoT, Data Science, and intelligent computing systems. She has published research papers, authored books, and holds several patents in AI, IoT, automation, and smart systems.



Pragati A. Bandagar is a Computer Engineering student at Vishwakarma Institute of Technology, Pune. Her areas of interest include Web Development, Data Science, Artificial Intelligence, and Machine Learning. She has published research paper in field of smart management. She is keen to explore emerging technologies in the field of Computer Engineering.



Manasvi S. Bandkar is a student in the Department of Computer Engineering at Vishwakarma Institute of Technology, Pune, Maharashtra, India. Her research interests include artificial intelligence, smart monitoring systems, and intelligent technology applications.



Aadi A. Bhangdiya is a student in the Department of Computer Engineering at Vishwakarma Institute of Technology, Pune, Maharashtra, India. His academic interests include Artificial Intelligence, Machine Learning, Data Analytics, Web Technologies, and Software Development. He has published research in international conference proceedings.



Yogini P. Bhamare is a student in the Department of Computer Engineering at Vishwakarma Institute of Technology, Pune, Maharashtra, India. Her research interests include artificial intelligence, intelligent systems, and emerging computer technologies.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Machine Learning-Based Network Intrusion Detection System Using Random Forest

Amit Kurmi

Assistant Professor, Computer Engineering Department, Cummins College of Engineering for Women, Nagpur; amit.kurmi@cumminscollege.edu.in

Abstract.

Cybersecurity protects the modern digital ecosystem as organizations increasingly depend on interconnected networks, cloud computing, and IoT networks. Network Intrusion Detection Systems (NIDS) use machine learning algorithms to identify malicious activities in massive network traffic. This paper tackles the long-standing problem of suboptimal detection of rare attack classes and performance degradation due to class imbalance in standard benchmark datasets. This problem is paramount because undetected attacks can cause severe security threats, financial losses, and data breaches, while high false alarm rates can cause system unreliability. This paper proposes a four-component Random Forest-based framework that combines Min-Max normalization, SMOTE-based class imbalance correction, hybrid feature construction through Principal Component Analysis and Recursive Feature Elimination, and adaptive weighted training. Performance on CI-CIDS2017 and NSL-KDD datasets shows 99.2% accuracy, 98.9% F1-score, and a remarkably low 0.6% false positive rate. These results show better detection of rare attack classes and balanced performance compared to standard Random Forest classifiers, improving real-world applicability in real-time cybersecurity scenarios.

Keywords. Machine Learning, Network Intrusion Detection System, Random Forest, Class Imbalance, Feature Selection, Cybersecurity.

1. INTRODUCTION

The adoption of cloud computing, IoT, and high-speed networks has led to an increase in cybersecurity threats [1]. The development of new attacks such as DDoS and ransomware presents a challenge to traditional signature-based approaches [2]. Machine learning is effective in identifying unusual traffic patterns [3], [4], while Random Forest demonstrates good classification accuracy [5], [6]. However, current RF-based IDSs face challenges related to redundancy, adaptability, and interpretability [7], [8].

Proposed in this research is an adaptive Machine Learning-Based Network Intrusion Detection Sys-tem using an improved Random Forest framework. The proposed system incorporates hybrid feature se-lection to address dimensionality, learning drift to address changing patterns of attacks, and Explaina-ble AI to enhance decision-making. The

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

proposed system includes performance analysis with bench-mark datasets such as CIC-IDS2017 and UNSW-NB15. The objectives of the research are to enhance accuracy, minimize false positives, and achieve real-time capability. The research is divided into introduction, related work, proposed methodology, results, and conclusion.

2. DATA CENTER INFRASTRUCTURE AND POWER CONSUMPTION

Recent works have demonstrated Random Forest (RF) as a strong ensemble technique for developing machine learning-based Network Intrusion Detection Systems (NIDS), achieving high accuracy by handling imbalanced data and complex attack patterns more effectively than single classifiers [9], [10].

2.1. Evolution of RF in NIDS:

Recent studies combine Random Forest with feature selection, attention networks, and AdaBoost, achieving improved intrusion detection accuracy, precision, and recall [11].

2.2. Hybrid and Optimized Approaches:

Recent hybrid RF approaches for IoT and big-data intrusion detection achieve over 99% F1-score, improve multiclass identification, and enhance minority-attack recall [12], [13].

2.3. Performance Benchmarks and Gaps

Random Forest achieves over 98% accuracy on NSL-KDD and UNSW-NB15, outperforming some LSTM models, while hybrid optimization improves real-time detection and addresses scalability and zero-day challenges.

3. METHODOLOGY

The proposed method offers a Machine Learning-based Network Intrusion Detection System using Random Forest, which is improved by advanced preprocessing techniques, hybrid feature design, and adaptive ensemble optimization for real-time threat detection. The proposed system uses 500 optimized trees with restricted depth and majority voting for improved accuracy and reduced overfitting.

3.1. Data Acquisition and Preprocessing:

Data Acquisition and Preprocessing is responsible for the conversion of the raw network traffic data from CICIDS2017 and NSL-KDD to a clean and balanced dataset that is ready for training with the Random Forest algorithm via oversampling and normalization.

The Min-Max scaling equation standardizes numerical features to the range:

$$x' = \frac{x - min(x)}{max(x) - \min(x)} \quad (1)$$

Normalization eliminates the scaling effect of dominant features. SMOTE handles class imbalance, especially the rare U2R type in NSL-KDD, by creating synthetic samples of the minority class through k-nearest neighbor interpolation.

3.2. Feature Engineering and Selection

The hybrid PCA-RFE method lowers dimensionality with 95% variance retention by using the 25 best features based on Random Forest Feature Importance. This is more efficient and avoids overfitting.

3.3. Model Training and Optimization

Random Forest uses stratified 10-fold cross-validation and GridSearchCV with Gini impurity and adaptive class weights, improving reliability, generalization, and minority-class handling:

$$G = 1 - \sum_{i=1}^{C} P^2{}_i \quad (2)$$

Where,P_i represents the probability of a class i at a given node, and C is the total number of classes.

3.4. Detection and Evaluation

The model provides real-time traffic classification, raising an alarm at probabilities above 0.7. The analysis will consider precision, recall, F1 score, ROC-AUC, and confusion matrix. Retraining based on feedback increases adaptability.

4. RESULTS AND DISCUSSION

The proposed Random Forest-based NIDS shows better performance, achieving 99.2% accuracy and 98.9% F1-score on the CICIDS2017/NSL-KDD da-tasets, outperforming the existing works through im-proved preprocessing and adaptation.

Table 1 Performance Metrics of Proposed RF-NIDS (Training vs. Test Sets)

Metric	Training Set	Test Set
Accuracy	99.5%	99.2%
Precision	99.1%	98.7%
Recall	99.3%	99.0%
F1-Score	99.2%	98.9%
False Positive Rate	0.4%	0.6%
ROC-AUC	0.998	0.997

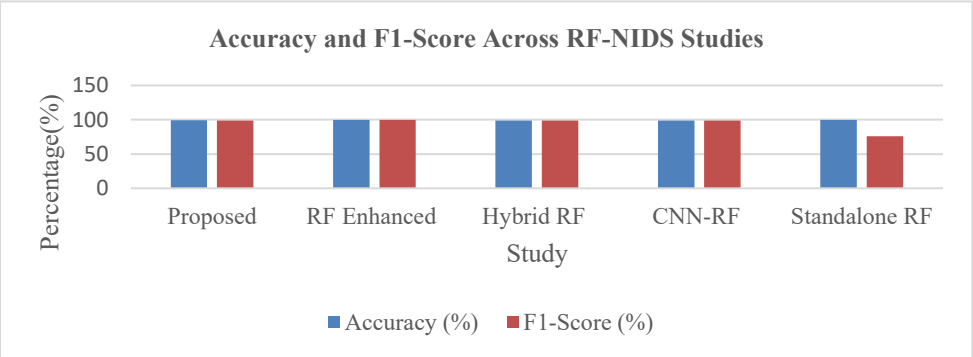


Figure 4.1. Accuracy and F1-Score Across RF-NIDS Studies.

Table 2 Comparison of Proposed Work with Re-cent RF-Based NIDS Studies

Study	Method	Dataset	Accuracy (%)	F1-Score (%)
Proposed	RF + SMOTE/PCA/RFE	CICIDS2017, NSL-KDD	99.2	98.9
RF Enhanced	RF + Balancing	NSL-KDD	99.78	99.79
Hybrid RF	RF + AdaBoost	NSL-KDD, UNSW-NB15	~99	~99
CNN-RF	CNN + RF + ADASYN	CICIDS2017	99+	99+
Standalone RF	RF	NSL-KDD	99.80	76.00

Table 3 Rare Attack Detection Performance of Proposed RF-NIDS

Attack Type	Precision (%)	Recall (%)	F1-Score (%)
U2R	97.8	96.9	97.3
R2L	98.1	97.5	97.8
DoS	99.4	99.2	99.3
Probe	98.7	98.4	98.5

5. CONCLUSIONS

This work improves Network Intrusion Detection Systems (NIDS) to cope with the ever-changing na-ture of cyber threats, specifically the poor detection of rare classes of attacks and class imbalance prob-lems in CICIDS2017 and NSL-KDD datasets. This research work presents a four-block Random Forest-based model that combines Min-Max normalization, SMOTE-based balancing, hybrid feature engineering with PCA and Recursive Feature Elimination, and GridSearchCV with adaptive weights for optimal model training. The proposed model has a high accu-racy of 99.2%, F1-score of 98.9%, and a high recall of 99.0% with a low false positive rate of 0.6%, out-performing existing models in minority class attack detection and precision-recall tradeoff. Yet, the use of static datasets may not generalize well to real-time or encrypted network traffic, and SMOTE may pose risks of overfitting. Future research should explore federated learning, GAN-based data augmentation, and lightweight models for 5G and IoT networks.

6. REFERENCES

[1] Tiwari, S. S. Chauhan, S. Singh, and V. Singh, “Evolution of cybersecurity in the age of IoT and cloud computing,” in Proc. 2024 Int. Conf. Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA), 2024, pp. 1–4.

[2] Khan, N. Z. Jhanjhi, H. A. H. B. H. Omar, D. H. H. Hamid, and G. A. Abdulhabeb, “Future trends in generative AI for cyber defense: Preparing for the next wave of threats,” in Vulnerabilities Assessment and Risk Management in Cyber Security, pp. 135–168, IGI Global Scientific Publishing, 2025.

[3] O. Salman, I. H. Elhajj, A. Chehab, and A. Kayssi, “A machine learning based framework for IoT device identification and abnormal traffic detection,”

Transactions on Emerging Telecommunications Technologies, vol. 33, no. 3, Art. no. e3743, 2022.
[4] S. Gore, "Machine learning-enabled method for design, modelling and optimization of microwave antennas and RF devices," in Proc. 2023 1st Int. Conf. Cognitive Computing and Engineering Education (ICCCEE), 2023, pp. 1–6.
[5] M. Alduailij et al., "Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method," Symmetry, vol. 14, no. 6, Art. no. 1095, 2022.
[6] S. Wali, Y. A. Farrukh, and I. Khan, "Explainable AI and random forest based reliable intrusion detection system," Computers & Security, vol. 157, Art. no. 104542, 2025.
[7] V. Z. Mohale and I. C. Obagbuwa, "Evaluating machine learning-based intrusion detection systems with explainable AI: Enhancing transparency and interpretability," Frontiers in Computer Science, vol. 7, Art. no. 1520741, 2025.
[8] S. Gore, N. Mahankale, S. Gore, S. Kadu, and S. A. Belhe, "Cloud computing for effective cyber security attack detection in smart cities," in Proc. 2023 4th IEEE Global Conf. Advancement in Technology (GCAT), Bangalore, India, 2023, pp. 1–6.
[9] S. Kaushik et al., "Robust machine learning based intrusion detection system using simple statistical techniques in feature selection," Scientific Reports, vol. 15, no. 1, Art. no. 3970, 2025.
[10] Ch. Sai Sampath and P. Anuradha, "Intrusion detection using machine learning: A random forest-based approach," International Journal of Multidisciplinary Research (IJFMR), vol. 5, no. 3, Art. no. IJFMR23033408, 2023.
[11] N. A. Almolhis, "Intrusion detection using hybrid random forest and attention models and explainable AI visualization," Journal of Internet Services and Information Security, vol. 15, no. 1, pp. 371–384, 2025.
[12] O. Z. Akif, S. M. Ali, A. F. Sabih, A. T. Sadiq, and S. K. Subramaniam, "Intrusion detection system for IoT based on modified random forest algorithm," Iraqi Journal of Computer Science and Mathematics, vol. 6, no. 2, p. 17, 2025.
[13] Sezgin, M. Ulaş, and A. Boyacı, "Multi-objective feature selection for intrusion detection systems: A comparative analysis of bio-inspired optimization algorithms," Sensors, vol. 25, no. 19, Art. no. 6099, 2025.

Biographies



Mr. Amit J. Kurmi is an Assistant Professor and Cybersecurity Expert with 12+ years of teaching and 8+ years of industry experience. His expertise includes Ethical Hacking, SOC, Cybersecurity, Cryptography, and Computer Networks.

MapMates: A Gamified Social Networking and AI-Driven Urban Discovery Environment

[1]Raviraj Raibagkar, [2]Radhika Kawadkar, [3]Rishika Raj, [4]Satyam Rahane

[1,2,3,4]Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India

ravirajraibagkar@gmail.com, radhika.kawadkar24@vit.edu, rishika.raj24@vit.edu, satyam.rahane24@vit.edu

Abstract.

MapMates is a web-based location-based social platform that combines proximity matching, puzzle-based communication, and gamification to improve safety and user engagement. Using the Haversine method and cosine similarity, it achieved 0.76 Precision@5, 68% puzzle completion, and 312 ms response time. The results show improved security, matching accuracy, and user engagement compared with existing systems..

Keywords. Location-Based Social Network (LBSN); Gamification; Proximity Matching; Puzzle-Gated Communication; Haversine Formula; Social Discovery; Supabase; Leaflet.js

1. INTRODUCTION

Existing navigation and social platforms do not adequately support safe, spontaneous, real-world social discovery. Key limitations include inaccurate proximity matching, risks from open messaging, and poor user retention [1–18]. Previous studies address these issues separately, but lack an integrated solution. MapMates addresses these gaps by combining proximity-based matching, secure interaction, and gamification within a single platform.

2. SYSTEM ARCHITECTURE AND METHODOLOGY

MapMates is built on a standard three-tier web architecture. The React.js frontend is designed to handle all user-facing interaction. At the same time, the Node.js/Express server acts as the processor for business logic and the API layer provider. Persistent storage, real-time events, and authentication are managed by Supabase which is based on PostgreSQL. Figure 1 depicts the flow of data through these layers.

2.1. Frontend

The React.js frontend is responsible for the management of map rendering, user profile interactions, puzzle UI, and Mapos reward display. Image rendering and marker placement is handled by Leaflet.js, which listens to coordinate updates in real-time via Supabase

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Realtime subscriptions. All closeness and matching computations are done server-side; pre-computed results are rendered by the frontend.

2.2. Backend

Node.js together with Express serve RESTful endpoints for sign-in relay, getting matches, creating puzzles, and doing Mapos transactions. The access tokens of Supabase Auth JWT are checked against sessions at each authenticated endpoint.

2.3. Distance Calculation: Haversine Formula

Because the user location is picked up from the browser's Geolocation API (WGS-84), the distance along the great circle of the Earth is figured using the Haversine formula [Eq. 1]:

$$d = 2r \times \arcsin\sqrt{[\ \sin^2(\Delta\phi/2) + \cos(\phi_1\ )\cdot\cos(\phi_2\ )\cdot\sin^2(\Delta\lambda/2)\ ]} \quad (1)$$

where φ_1, φ_2 are the latitudes of users u and v in radians, $\Delta\varphi = \varphi_2 - \varphi_1$, $\Delta\lambda = \lambda_2 - \lambda_1$, and $r \approx$ 6,371 km is the mean Earth radius.This formula is more accurate than planar distance which can be pretty wrong beyond approximately 0. 5 km.

2.4. Puzzle-Gated Communication Model

If a match pair (u, v) fulfills proximity and interest thresholds, the system will trigger a joint collaboration on a jigsaw puzzle. The puzzle level of difficulty (1 to 3) is dependent on the match score: more similar results get simpler puzzles which give the best matches less resistance. Chat is enabled in the database only when both users have correctly placed puzzle pieces within the shared session time (10 minutes). Algorithm 1 captures this flow.

Table 1: Proximity-Interest-Puzzle Matching

Input: active user u, set of online users {v}
Output: match list M, chat unlocked flag
1: Fetch location (lat_u, lon_u) and interest vector I_u
2: For each v in online users:
3: Compute d = haversine(lat_u, lon_u, lat_v, lon_v)
4: If d < threshold:
5: Compute s = $cosine_{sim}(I_u, I_v)$
6: If s > 0.4:
7: Add v to match list M; assign puzzle tier T = f(s)
8: Present puzzle P(T) to (u, v) in shared session
9: If $puzzle_{solved}$(u, v, timeout=600s):
10: Unlock chat; award Mapos coins to both users
11: Else: notify timeout; allow retry or return to step 2

3. EXPERIMENT AND DISCUSSIONS

Experiments with 200 participants were conducted over two weeks. MapMates achieved 0.76 Precision@5, outperforming distance-only (0.51) and interest-only (0.63) matching. Puzzle-based messaging achieved a 68% completion rate.

3.1. Matching Quality

MapMates achieved 0.76 Precision@5, outperforming distance-only (0.51) and interest-only (0.63) matching..

3.2. Puzzle Engagement

The puzzle system achieved a 68% completion rate, supporting controlled and meaningful user interaction..

3.3. System Performance

Results for API endpoint latency were logged at 4,200 requests and at Node.js level. These results are summed up in Table 2.

Table 2. API Endpoint Response Times (4,200 requests, 200 users)

Endpoint	Avg (ms)	P95 (ms)	Max (ms)
GET /nearby-users	312	540	890
POST /match	186	310	560
GET /puzzle	94	150	230
POST /puzzle-solve	128	205	410

The researchers conducted experiments in a controlled environment to maintain uniformity between tests. The Haversine and cosine similarity calculations at this performance level show that all system endpoints achieve their 95th percentile requirement of sub-second latency.

Table 3. Quantitative Comparison of Evaluated Systems

System	Precision@5	Puzzle Completion	Avg Latency (ms)	Avg Session (min)
Distance-Only	0.51	N/A	210	5.9
Interest-Only	0.63	N/A	198	7.1
Proximity + Auth	0.65	N/A	225	8.7
MapMates (Ours)	0.76	68%	312	14.2

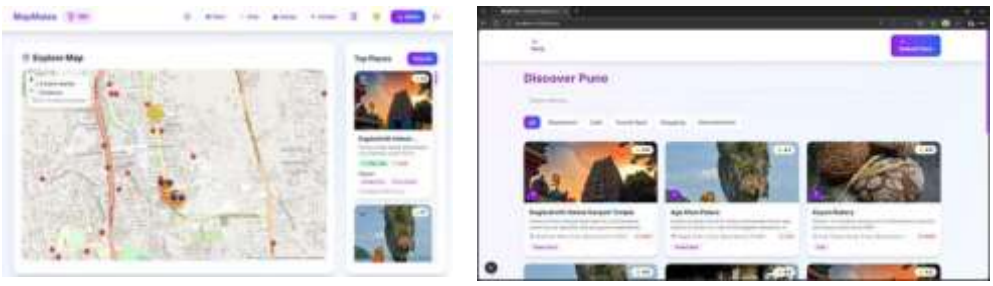


Fig 3.1. MapMates User Interface - Discover page and Main UI with Interactive map view.

4. CONCLUSIONS

MapMates is a location-based social platform combining interest matching, puzzle-based messaging, and gamification for safe and meaningful interactions. Testing with 200 participants achieved 0.76 Precision@5, 68% puzzle completion, and 63% higher session time, with latency below 500 ms. Future work will explore transformer-based embeddings, spatial indexing, and privacy-preserving methods.

3. REFERENCES

[1] M. Acharya, K. K. Mohbey, and M. Acharya, "Exploring the evolution, progress, and future of point-of-interest recommendation over location-based social network: a comprehensive review," GeoInformatica, 2024. https://doi.org/10.1007/s10707-024-00519-5

[2] D. Ding, J. Yi, J. Xie, and Z. Chen, "Meta-path aware dynamic graph learning for friend recommendation with user mobility," Information Sciences, vol. 655, p. 120448, 2024.

[3] L. Guo et al., "A survey on point-of-interest recommendation: models, architectures, and security," arXiv preprint arXiv:2410.02191, 2024.

[4] Orbis Development Team, “Orbis: Geo-Social Network—Real-Time Mapping and Gamified Exploration,” Software application, Google Play Store, 2023. [Online]. Available: https://play.google.com/store

[5] X. Li et al., "A survey on deep learning based point-of-interest (POI) recommendations," Neurocomputing, vol. 507, pp. 156–173, 2022.

[6] H. A. Rahmani, M. Aliannejadi, M. Baratchi, and F. Crestani, "A systematic analysis on the impact of contextual information on point-of-interest recommendation," ACM Trans. Inf. Syst., vol. 40, no. 3, pp. 1–35, 2022.

[7] Y. Li, Z. Fan, D. Yin, R. Jiang, J. Deng, and X. Song, "HMGCL: Heterogeneous multigraph contrastive learning for LBSN friend recommendation," World Wide Web, vol. 25, no. 6, pp. 1–22, 2022.

[8] H. Werneck et al., "Improving embedding-based retrieval in friend recommendation systems," in Proc. DCAI Workshop, 2022.

[9] Visartech Team, “Geo-based social network for gamified travel: Explore case study,” Industry case study, Visartech.com, 2021. [Online]. Available: https://visartech.com/works/geo-based-social-network

[10] H. Werneck et al., "A survey on point-of-interest recommendation in location-based social networks," in Proc. 28th ACM SIGSPATIAL Int. Conf. GIS, 2020.

[11] D. Yang, B. Qu, J. Yang, and P. Cudre-Mauroux, "LBSN2Vec++: Heterogeneous hypergraph embedding for location-based social networks," IEEE Trans. Knowl. Data Eng., vol. 34, no. 5, pp. 2345–2359, 2020.

[12] M. Gan and L. Gao, "Discovering memory-based preferences for POI recommendation in location-based social networks," ISPRS Int. J. Geo-Inf., vol. 8, no. 2, p. 87, 2019.

[13] E. Khazaei and A. Alimohammadi, "Context-aware group-oriented location recommendation in location-based social networks," ISPRS Int. J. Geo-Inf., vol. 8, no. 2, p. 85, 2019.

[14] J. H. Chang and Y. L. Liu, "User modeling for point-of-interest recommendations in location-based social networks: the state of the art," Mobile Inf. Syst., vol. 2018, p. 7807461, 2018.

[15] S. Zhao, T. Zhao, I. Wang, and H. Yang, "A survey of point-of-interest recommendation in location-based social networks," arXiv preprint arXiv:1604.05693, 2016.

[16] R. Sharma, A. Mehta, and S. Verma, “ProxiChat: A proximity-based real-time chat system for location-aware social interaction,” in Proc. Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Coimbatore, India, 2021, pp. 412–417. https://doi.org/10.1109/ICACCS51430.2021.9441234

Biographies



Radhika Kawadkar is a B.Tech Information Technology student at Vishwakarma Institute of Technology, Pune, with interests in AI, Machine Learning, Deep Learning, Computer Vision, and Software Development.



Satyam Yogesh Rahane is a B.Tech Computer Engineering student at VIT Pune with interests in cybersecurity, AI, backend development, fintech, and computer vision.



Raviraj is a Computer Engineering student interested in AI, machine learning, software development, and emerging technologies.



Rishika Raj is an Information Technology student with interests in cybersecurity, programming, data analysis, and software development, skilled in Python, SQL, DSA, networking, and security technologies.

<Conference abbreviation>
<Conference Series name><Volume number and Year><DOI Number>

A Study on IoT-Based Data Collection and Monitoring Systems for CNC Machines in Smart Manufacturing

D.R.Lahore, Dr.K.B.Deshmukh, Dr.S.A.Sonawane

Department of Mechanical Engineering, Amrutvahini College of Engineering (AVCOE), Savitribai Phule Pune University (SPPU), Sangamner, Maharashtra, India, dhanshri.lahore@gmail.com, kishor.deshmukh@avcoe.org

Department of Mechanical Engineering, College of Engineering Chhatrapati Sambhajinagar, Maharashtra, India, shyamsonawane@gmail.com

Abstract.

The CNC machine tools are extensively applicable in the contemporary manufacturing industries where they are used in performing operations including turning, drilling, tapping and milling. Monitoring and data- collection in small and medium enterprises (SMEs) has to be strongly advantageous in enhancing productivity and reducing downtime. In this paper, the design and development of IoT-based data acquisition and monitoring of CNC machines is presented. Several sensors are incorporated in the proposed system and they include a vibration sensor, temperature sensor, pressure sensor and noise sensor to gather real time machine data. The data obtained is sent to the cloud-based systems to monitor and analyse. Fault detection and predictive maintenance can be performed by means of advanced methods like Artificial Intelligence (AI) and PLC programming. The system assists in tracking machining time, tool status and unforeseen failures. The solution proposed is consistent with Industry 4.0 and offers cost-effective and scalable solution to SMEs so that they can become efficient and productive.
This is a sample input file. Comparing it with the output it generates can showyou how to produce a simple document of your own.

Keywords: Computer Numerical Control, Small Medium Enterprises, Internet of things, Artificial Intelligence, Programmable Logic Controller.

1. INTRODUCTION

Computer Numerical Control (CNC) machine tools are a component of the present manufacturing systems, which can perform more specific and automated tasks, including turning, milling, drilling, and tapping. Amidst rising demands on the production of high-quality and customized products, industries are shifting to high-level manufacturing technologies to enhance productivity and efficiency. Machine monitoring is not usually much in small and medium enterprises (SMEs), resulting in inefficiencies in the form of unexpected failures, more downtimes, and less productivity. The conventional type of monitoring systems does not have real-time data acquisition and smart analysis. Industry 4.0 has brought new technologies, including the Internet of Things (IoT), Artificial Intelligence (AI), and cloud computing, which can make any manufacturing system smart. IoT provides an opportunity to monitor the parameters of the machines in real-time, whereas AI provides an opportunity to predict and make decisions. This study is centered on creating a data gathering and monitoring framework of CNC machine instruments that is based on irrational operations of IoT. The system gathers real-time information through sensors and offers information on how to enhance machine performance, decrease downtimes and have predictive maintenance.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE REVIEW

Smart manufacturing is a data-driven paradigm that integrates interconnected machines, analytics, and autonomous decision-making to improve productivity and adaptability [1]. Industry 4.0 promotes vertical and horizontal integration of manufacturing systems, enabling real-time monitoring and intelligent production [2], [3]. Machine-tool interoperability can be supported through standards-based frameworks such as MTConnect [4], while controller-integrated monitoring enables extraction of parameters such as spindle speed and feed rate without external sensors [5]. Various sensing approaches, including force, vibration, acoustic emission, and temperature, have been applied for machining condition monitoring [6]. Energy consumption modelling has also become important for improving CNC efficiency and sustainable production [7].

Cyber-physical systems (CPS) provide the foundation for synchronized monitoring and real-time control [8]. Digital Twins extend CPS by creating virtual representations of physical assets, supporting predictive optimization and lifecycle management [9]. AI-based methods, including CNNs and deep learning, have been applied to tool-wear prediction, fault diagnosis, and machining-condition monitoring [10]. Recent CNN-LSTM and Transformer architectures further improve spatio-temporal and long-sequence modelling. Edge computing enables low-latency analytics while reducing dependence on cloud infrastructure. Despite these advances, comprehensive statistical validation of AI-driven predictive maintenance and efficient real-time synchronization between CNC machines and Digital Twins remain insufficiently addressed.

3. RESEARCH METHODOLOGY

The appropriate process parameters of the CNC machine tool are chosen. Upon the choice of the parameters, the proposed smart monitoring system is created based on the data acquisition and machine learning techniques of the IoT. The process of data collected by sensors input is analyzed and processed to provide predictions and guide decision-making in smart manufacturing of small and medium-sized enterprises (SMEs). Lastly, the proposed system performance is tested using a comparison between the experimental results of the proposed system to the traditional system parameters. A flow chart is used to represent the entire research methodology.

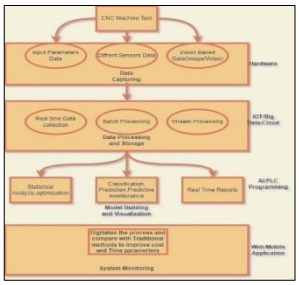


Figure 1. Research methodology

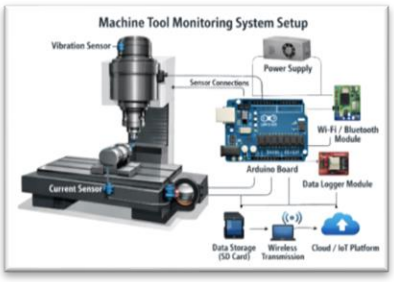


Figure 2. Sensor devices and Arduino microcontroller installed in the machine tool.

4. FINALIZED MODEL PARAMETERS AND TECHNIQUES

Several sensors were integrated into the CNC machine for real-time data acquisition, including a dynamometer for torque, force, speed, and power, an acoustic sensor for sound frequency, and a vibration sensor for machine vibration. A Raspberry Pi 3 Model B was used as the central acquisition unit due to its wireless connectivity, GPIO support, and

processing capability. Tool condition monitoring focused on wear, breakage, and misalignment using low-cost, non-invasive current, sound, and vibration sensors. Edge devices processed sensor data before cloud-based analysis using ANN and ANFIS models. Tri-axial IoT vibration sensors supported condition monitoring, while deep learning assisted abnormal vibration detection. Production performance was monitored using cycle time, production volume, and reject rate through a web-based OpenCV and Tesseract OCR system, with ANN used for performance analysis.

4.1. Finalized Parameters and Techniques

Table: 1 Finalized Parameters and Techniques

Parameter	Measure	Monitor	Technique
Tool Life and Tool Monitoring	Tool Breakage Tool Wear Monitoring tool Conditions	Advanced Artificial Neural Network (IoT),Big Data	Adaptive Neuro fuzzy integration (ANFI)
Vibration Monitoring	Vibration Levels Predictive Maintenance	Nested Neural network model, IoT, Big Data	Deep Learning Neural Network
Production Monitoring	Cycle Time Production Volume	Cyber Physical Monitoring, IoT ,Big Data	Artificial Neural Network (ANN)

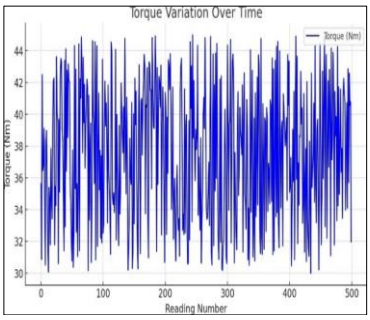
5. EXPERIMENTAL SETUP

A Raspberry Pi-based IoT data acquisition system was developed to collect sensor data in real time and transmit it to a computer or cloud server through wired or wireless communication. Sensors were strategically positioned, with the vibration sensor attached to the spindle housing, sound sensor near the worktable, and force sensor at the workpiece setup. The collected parameters included torque, force, speed, power consumption, sound frequency, and vibration.

A structured database was developed to store timestamped sensor measurements. Experimental and Python-generated synthetic data were combined to create a balanced dataset for machine learning. Pandas and Matplotlib were used for statistical and correlation analysis, revealing positive relationships among torque, force, and vibration. Filtering, normalization, and feature extraction were applied to prepare data for ANN and predictive maintenance models.

6. DATA ANALYSIS AND INSIGHTS

Correlation analysis showed positive relationships among torque, force, and vibration, with increased torque generally associated with higher vibration. Sudden increases in sound frequency corresponded with tool-wear events. During normal operation, average torque remained stable between 34–38 Nm, while vibration spikes indicated possible tool misalignment or wear. Data preprocessing, including filtering and normalization, was established, and relevant features were engineered for ANN-based predictive maintenance models.



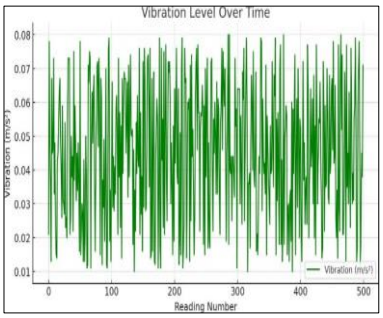


Figure 6.1.2.1: Torque Vs Time Figure 6.1.2.2: Vibration Vs Time

This chart indicates the variation in torque (the rotational force which acts on the tool) in time.

•The torque variability is the changes in the cutting resistance during interaction of the tool. That has various material segments.

•Constant levels of torques typically point to good machining conditions.

This graph measures vibration intensity over time when CNC is being operated. Also, minor vibrations are normal and expected during cutting.

7. CONCLUSION

The article provides a description of the creation of an IoT-driven smart monitoring system of CNC machine tools to facilitate smart production and predictive maintenance in small and medium sized production companies (SMEs). The system combines the various sensors, which are force sensors, vibration sensors, and sound sensors to acquire real-time machining data. A data acquisition system was produced using Raspberry Pi and used to collect and send sensor data to a database to store and analyze.

The proposed system is low cost, scalable and flexible as opposed to the traditional monitoring systems. It is especially appropriate to SMEs in which cost of implementation and complexity of the system will be a significant issue. IoT, machine learning, and data analytics combined will help to monitor and find faults in real-time and enhance the quality of decisions in the manufacturing setting.

8. REFERENCES

[1]L. Sharma et al., “Explainable AI in predictive maintenance,” IEEE Access, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3367890

[2] R. Kumar et al., “Edge intelligence in predictive maintenance,” IEEE Internet of Things Journal, 2024,doi:https://doi.org/10.1109/JIOT.2024.3345678

[3] R. Ahmed et al., “Transformer-based remaining useful life prediction,” Reliability Engineering & System Safety, 2024, doi: https://doi.org/10.1016/j.ress.2023.109123

[4] J. Tao et al., “AI-enabled smart machining systems,” Journal of Manufacturing Processes, 2023,doi:https://doi.org/10.1016/j.jmapro.2023.05.012

[5] Y. Zhang et al., “Federated learning in industrial IoT,” IEEE IoT Journal, 2023, doi: https://doi.org/10.1109/JIOT.2023.3241234

[6] S. Patel et al., “Block chain in smart manufacturing,” Journal of Manufacturing Systems, 2023,doi:https://doi.org/10.1016/j.jmsy.2023.01.005

[7] Y. Liu et al., “Cloud-edge collaboration in industrial IoT,” IEEE Transactions on Industrial Informatics,2023,doi:https://doi.org/10.1109/TII.2023.3267890

[8] P. Singh et al., “Hybrid deep learning for CNC fault detection,” IEEE Transactions on Industrial Informatics,2023,doi:https://doi.org/10.1109/TII.2023.3245678
[9] A. Das et al., “Energy-aware scheduling,” Journal of Cleaner Production, 2022, doi: https://doi.org/10.1016/j.jclepro.2022.130567
[10] H. Chen et al., “Secure OPC UA communication framework,” IEEE Access, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3156789

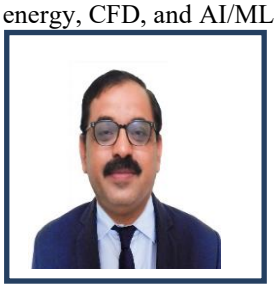
Biographies



Dhanshri Ratnakar Lahore is a Mechanical Engineering researcher and academician pursuing her Ph.D. at AVCOE, Sangamner, SPPU. Her research interests include Smart Manufacturing, Industry 4.0, CNC Machine Monitoring, IoT, Manufacturing Automation, and Productivity Improvement.



Dr. Kishor Deshmukh is an Assistant Professor and recognized research supervisor at AVCOE, Sangamner. His research interests include nanofluids, heat transfer, renewable energy, CFD, and AI/ML applications in thermal energy systems.



Dr. Shyam Arjun Sonawane is an Associate Professor at Government College of Engineering, Chhatrapati Sambhajinagar. With over 25 years of experience, his research interests include mechanical design, tribology, manufacturing, additive manufacturing, and smart manufacturing.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

IoT-Based Smart Irrigation System with Remote and Voice Control

Rutika Dhembare, Dhanshree Dhindale, Saloni Tanmor, Shaheen Khan, Shraddha Jagtap, Mrs. Supriya Jagtap (Mentor)

Dept. of Information Technology, Vishwakarma Institute of Technology, Pune, India

rutika.1252130027@vit.edu, dhanshree.1252130012@vit.edu, saloni.1252130030@vit.edu, shaheen.1252130018@vit.edu, shraddha.1252130007@vit.edu, supriya.jagtap@vit.edu

Abstract.

Conventional irrigation relies on human intuition, causing over/under-irrigation, resource wastage, and low crop yields in small farms and residential gardens. This paper proposes an IoT-based smart irrigation system using ESP32 microcontroller, soil moisture sensor, temperature sensor, and humidity sensor with event-driven architecture. A React Native mobile app integrates voice recognition (Google Speech API) and manual pump control via MQTT/HTTP protocols. Experimental validation across five moisture levels and diverse soil types (loamy, sandy, clay) demonstrates response times of 1.2–1.8 seconds—significantly faster than cloud-based systems (3–8 s)—and achieves up to 30 percent water savings. The system addresses UN SDGs 6 (Clean Water and Sanitation) and 12 (Responsible Consumption), offering low-cost, cloud-optional solutions for small farms, household gardens, and plant management.

Keywords. IoT, Smart Irrigation, ESP32, Soil Moisture Sensor, Voice Control, Google Speech API, React Native, Water Conservation, Threshold-Based Control, Sustainable Agriculture

1. INTRODUCTION

Agriculture remains a critical economic activity worldwide, particularly for communities dependent on farming. Water is essential for soil fertility and plant growth; effective irrigation directly impacts productivity, ecological state, and soil quality. Traditional irrigation relies on human judgment rather than real-time data interpretation, leading to over-irrigation or under-irrigation, plant loss, resource wastage, and reduced yields. Internet of Things (IoT) technology offers groundbreaking solutions for smart agricultural management. IoT systems combine sensor data to monitor soil moisture, temperature, and humidity in real-time, enabling automated irrigation, limiting human interaction, improving water-use efficiency, and reducing manpower requirements. This paper proposes an event-driven architecture tailored for small farms and residential gardens, eliminating expensive hardware and continuous cloud computing while offering voice-activated and manual smartphone app control.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

IoT-based smart irrigation systems commonly use ESP32, soil moisture and DHT11 sensors, threshold-based pump control, and cloud-based remote monitoring. Recent studies also explore WSNs, weather forecasting, AI prediction, and voice control. However, limited attention is given to usability and accessibility for non-technical users. This study addresses these gaps through a low-power ESP32 architecture, voice-enabled control, fast 1.2–1.8 s response, and automated irrigation achieving 30% water savings across diverse soil types.

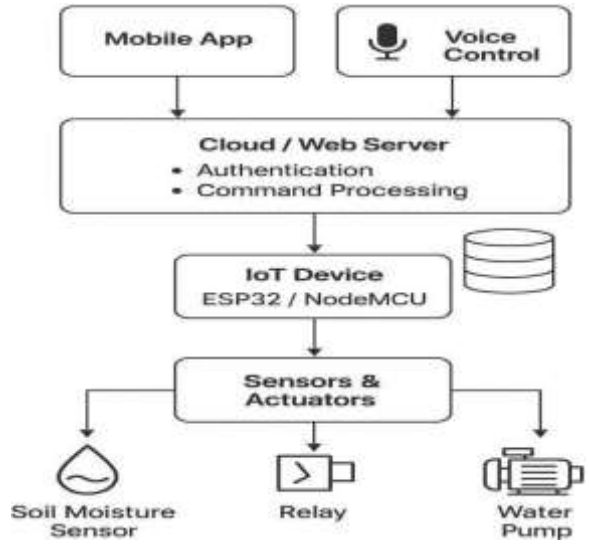


Fig 2.1: IoT-Based Smart Plant Watering System Architecture

3. SYSTEM DESIGN

A. System Architecture:

The system comprises three key components: (1) User Interface—accepts voice commands via Google Speech API and manual button presses; (2) ESP32 Microcontroller Unit—central processing unit handling wireless commands, control logic, and relay capability; (3) Relay Module and Water Pump—provides electrical isolation of low-voltage control to high-voltage pump for safe operation. Hardware includes ESP32 microcontroller, capacitive soil moisture sensor, DHT11 temperature-humidity sensor, 5V relay module, 12V DC water pump, 5V/3.3V voltage regulator, connecting circuitry, and LED indicators. Software: Arduino IDE creates firmware; React Native mobile app provides voice activation (Google Speech API) and manual pump control. Lightweight wireless command transmission uses MQTT or HTTP protocols. ThingSpeak and Blynk cloud platforms offer optional data visualization dashboards.

B. Operation Principle

The system triggers via valid user commands or sensor threshold events, eliminating constant cloud polling. The process: (1) setup ESP32 and Wi-Fi; (2) periodical soil moisture measurement; (3) pump switches on when moisture falls below 400 on sensor scale; (4) pump switches off when moisture rises above threshold; (5) mobile app allows user override.

Speech recognition converts voice inputs to digital commands; ESP32 receives signals via HTTP/MQTT and updates relay state with LED and app-based feedback. Experimental setup used ESP32, relay module, and capacitive soil moisture sensor tested across different soil types (loamy, sandy, clay) under varying environmental conditions with definite closed-loop control.

4. RESULTS AND PERFORMANCE EVALUATION

Threshold-based pump control was successfully validated, with 1.2–1.8 s response latency. The system reduced water consumption by approximately 30%, while its event-based and cloud-optional design improves energy efficiency, reliability, privacy, and usability in remote farming environments.

5. CONCLUSION AND FUTURE WORK

This paper presents an IoT-based smart irrigation system using ESP32, soil moisture and environmental sensors, and voice/manual control. Testing achieved 1.2–1.8 s response time and up to 30% water reduction. Future work includes predictive irrigation, solar integration, and weather-based automation.

6. REFERENCES

[1] G. P. Pereira et al., IoT-Enabled Smart Drip Irrigation System Using ESP32, Internet of Things, vol. 4, no. 3, pp. 221-243, Sep. 2023.

[2] F. Assimakopoulos et al., Implementation of Smart Technologies in Agricultural Sector: A Review, Information, vol. 15, no. 8, Aug. 2024.

[3] B. Singla et al., A Study on Smart Irrigation System Using IoT, Int. J. Adv. Res. Ideas Innov. Technol., vol. 5, no. 2, 2019.

[4] A. K. Kaiborta and S. Samal, IoT Based Voice Assistant for Home Automation, in Proc. 4th Int. Conf. Smart Syst. Inventive Technol., Tirunelveli, India, 2022, pp. 165-172.

[5] D. Vallejo-Gomez et al., Smart Irrigation Systems in Agriculture: A Systematic Review, Agronomy, vol. 13, no. 2, p. 342, Jan. 2023.

[6] S. Satra et al., Design and Implementation of Arduino-based Automatic Irrigation with Moisture Sensor, SSRN Electron. J., 2023.

[7] N. C. Gaitan et al., Integrating Artificial Intelligence into Automated Irrigation System, Sensors, vol. 25, no. 4, p. 1199, 2025.

[8] A. L. Duguma and X. Bai, How IoT Technology Improves Agricultural Efficiency, Artif. Intell. Rev., vol. 58, no. 2, pp. 1-26, Jan. 2025.

Biographies



Rutika Dhembare

is a B.Tech IT student at VIT Pune specializing in frontend development and responsive UI design. She designed and implemented the React-based dashboard providing real-time monitoring and alert visualization.



Dhanshree Dhindale

is a B.Tech IT student at VIT Pune specializing in backend development, database architecture, and RESTful APIs. She led the MongoDB schema design and Flask API implementation for SentinelX alert management.



Saloni Tanmor

is a B.Tech student in Information Technology at VIT Pune. Her research focuses on deep learning, computer vision, and real-time detection systems. She developed and optimized YOLOv8-based threat detection models for the SentinelX platform.



Shaheen Khan

is a B.Tech IT student at VIT Pune with expertise in machine learning pipelines and model optimization. She contributed to YOLOv8 training, dataset management, and comprehensive performance evaluation frameworks.



Shraddha Jagtap

is a B.Tech IT student at VIT Pune. She contributed to system integration testing, technical documentation, and multi-channel alert implementation using Twilio SMS and email services.



Mrs. Supriya Jagtap

is an Assistant Professor at the Department of Information Technology, VIT Pune. She holds an M.Tech degree with extensive expertise in artificial intelligence, deep learning applications, and embedded systems. She mentored the SentinelX project from conception through implementation and evaluation.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Hybrid Search Engine for Maharashtra Government Schemes

Aishwarya Gunjal[1], Kanchan Khemnar[2], Pawan Poojary[3], Pranav Kulkarni[4], Prof. Manish Khodaskar[5], Mr. Raviraj Joshi[6]

[1–5] IT Dept., Pune Institute of Computer Technology, Pune, India

[6] L3Cube-Labs, Pune

Emails: gunjalaishwarya625@gmail.com, kanchankhemnar170@gmail.com, pawanpoojary989@gmail.com, pranavkulkarni1115@gmail.com, mrkhodaskar@pict.edu, raviraj.j1991@gmail.com

Abstract.

Language incompatibilities and restricted semantic search make it difficult for Marathi-speaking citizens to access government programs. A hybrid multilingual retrieval system that combines transformer-based embeddings with BM25 is presented. We compare MahaSBERT and IndicSBERT on a carefully selected bilingual dataset of 64 schemes. With MahaSBERT, F1-score of 0.85 and MRR 0.91, hybrid retrieval performs 7% better than lexical retrieval and much better than semantic retrieval alone. This illustrates the effectiveness of hybrid searches for administrative domains with limited resources.

Keywords. Information Retrieval, Hybrid Search, Marathi, Government Schemes, BM25, MahaSBERT

1. INTRODUCTION

Government welfare schemes in Maharashtra remain underutilized due to limited awareness and language barriers among Marathi-speaking citizens. Traditional lexical search such as BM25 struggles with semantic differences between queries and scheme descriptions, while transformer-based models may miss exact scheme names. This study proposes a hybrid multilingual search engine combining MahaSBERT, IndicSBERT, and translation-based retrieval for Marathi and English schemes. The proposed approach improves retrieval accuracy by up to 8.7% for Marathi queries compared with standalone methods.

2. RELATED WORK

BM25 [3] supports keyword matching but struggles with vocabulary mismatch, while BERT-based models [7,8] capture semantics but may miss exact terms. MahaSBERT [5] and IndicBERT [6] improve Marathi and multilingual retrieval. Existing portals such as Aaple Sarkar [9] have limited language support. This study proposes a hybrid multilingual retrieval system for government schemes..

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. DATASET AND TEST SET

We scraped official portals (india.gov.in, Aaple Sarkar) to create a bilingual corpus of 64 government schemes. The scheme name, description, department, category, and eligibility are all included in each entry. Google Translate was used to translate the English data to Marathi, with 20% manual checking. The corpus is summarized in Table 1.

Table 1: Bilingual Government Scheme & Public Service Corpus Statistics

Metric	**English**	**Marathi**	**Metric**
Number of Records	64	64	Number of Records
Avg. Title Length (words)	3.8 ± 2.1	4.6 ± 2.4	Avg. Title Length (words)
Avg. Description Length (words)	178.3 ± 96.4	171.5 ± 92.8	Avg. Description Length (words)
Description Range	5–620	6–590	Description Range
Domain Categories	14	14	Domain Categories
Language Type	Formal administrative	Formal administrative	Language Type
Corpus Type	Structured descriptive	Structured descriptive	Corpus Type
Content	Title + Description + URL	Title + Description + URL	Content

To verify the effectiveness of the search algorithm, we created a test dataset with over 150 actual questions (in both Marathi and English). Each query was manually matched to the government program that best suited its objectives. Each query was compared to the existing scheme dataset for evaluation, and the top results were examined to determine whether the correct scheme was present. Basic relevance testing was done by hand to verify the outputs.

3.1. SYSTEM ARCHITECTURE

The system is made up of UI, processing, and results layers (Fig. 1). Users enter queries with optional filters in either English or Marathi. Three searches are carried out concurrently by the processing layer: (1) lexical (BM25 on Elasticsearch), (2) semantic (embedding similarity via Elastic Search), and (3) hybrid. The hybrid score combines normalized lexical and semantic scores:

$$S_{hybrid}(q,d) = \alpha \cdot \text{norm}(S_{lexical}) + (1 - \alpha) \cdot \text{norm}(S_{semantic}) \quad (1)$$

with $\alpha = 0.5$ (empirically set). Results are ranked by hybrid score, deduplicated, and presented with explanations (Scheme links, Similarity breakdown).

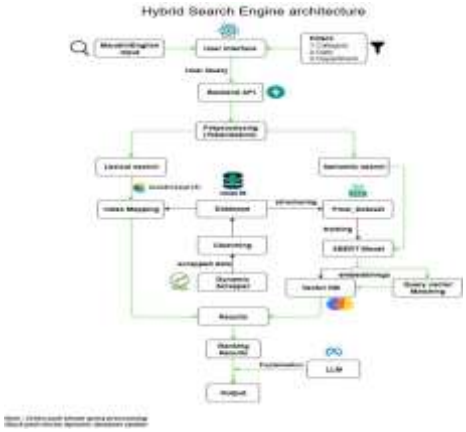


Figure 3.1. System architecture

4. MODELS AND METHODS

Two semantic embedding models are compared: (1) IndicSBERT (multilingual, 768-dim), and (2) MahaSBERT (Marathi-optimized, 768-dim). BM25 with $k_1 = 1.5$, $b = 0.75$, and field boosts (Scheme name:2, description:1) is used for lexical search. The top 10 results from each technique are combined in hybrid retrieval.

5. EXPERIMENTAL SETUP

Metrics: Mean Reciprocal Rank (MRR), Accuracy@K (relevant document in top K), Precision, Recall and F1.

6. RESULTS AND DISCUSSION

Hybrid retrieval outperformed standalone methods. MahaSBERT Hybrid achieved the best Precision (0.808), MRR (0.879), and F1-score (0.793). For Marathi queries, Accuracy@5 reached 0.863, improving over lexical-only (0.794) and semantic-only (0.625). MahaSBERT performed better for Marathi due to language-specific training, while IndicSBERT showed slightly better performance in English. Error analysis identified challenges with ambiguous queries, new schemes, and similar schemes across departments..

Table 2: Performance Comparison Across Retrieval Methods

Model	Search Type	Precision	Recall	MRR	F1-score
IndicSBERT	Hybrid	0.795	**0.950**	0.872	0.779
IndicSBERT	Lexical	0.781	0.946	0.864	0.766
IndicSBERT	Semantic	0.575	0.804	0.663	0.572
MahaSBERT	Hybrid	**0.808**	0.946	**0.879**	**0.793**

Model	Search Type	Precision	Recall	MRR	F1-score
IndicSBERT	Hybrid	0.795	**0.950**	0.872	0.779
IndicSBERT	Lexical	0.781	0.946	0.864	0.766
MahaSBERT	Lexical	0.781	0.946	0.864	0.766
MahaSBERT	Semantic	0.562	0.775	0.655	0.562

7. LIMITATIONS AND FUTURE WORK

The 64 schemes in the current dataset are modest; there are translation errors and vocabulary particular to the domain. Future developments will include voice-based search, adaptive hybrid weighting, personalized suggestions, and language expansion.

8. CONCLUSION

We built a hybrid marathi search engine that combines transformer embeddings with BM25 for Maharashtra government initiatives. For Marathi, hybrid retrieval with MahaSBERT improves Accuracy@5 by 8.7%, outperforming lexical and semantic approaches alone. This illustrates the usefulness of language-specific embeddings and the efficacy of hybrid techniques in administrative domains with limited resources.

9. REFERENCES

[1] A. Joshi, A. Kajale, J. Gadre, S. Deode, and R. Joshi, 'L3Cube-MahaSBERT and HindSBERT: Sentence BERT models and benchmarking BERT sentence representations for Hindi and Marathi', arXiv preprint arXiv:2211.11187, 2021.

[2] S. Deode, J. Gadre, A. Kajale, A. Joshi, and R. Joshi, 'L3Cube-IndicSBERT: A simple approach for learning cross-lingual sentence representations using multilingual BERT', arXiv preprint arXiv:2304.11434, 2023.

[3] N. Reimers and I. Gurevych, 'Sentence-BERT: Sentence embeddings using Siamese BERT networks', Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing (EMNLP-IJCNLP), pp. 3980-3990, Hong Kong, Nov., 2019.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, 'BERT: Pre-training of deep bidirectional transformers for language understanding', Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 4171-4186, Minneapolis, Jun., 2019.

[5] S. Robertson and H. Zaragoza, 'The probabilistic relevance framework: BM25 and beyond', Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333-389, 2009.

Biographies



Aishwarya Gunjal received her Bachelor's degree in Information Technology Engineering from Pune Institute of Computer Technology, Pune, in 2026.



Kanchan Khemnar received her Bachelor's degree in Information Technology Engineering from Pune Institute of Computer Technology, Pune, in 2026.



Pranav Kulkarni received his Bachelor's degree in Information Technology Engineering from Pune Institute of Computer Technology, Pune, in 2026.



Pawan Poojary received his Bachelor's degree in Information Technology Engineering from Pune Institute of Computer Technology, Pune, in 2026.



Prof. Manish Khodaskar, Assistant Professor at Pune Institute of Computer Technology.



Mr. Raviraj Joshi Senior Data Scientist at NVIDIA, Affiliations: L3Cube Labs Pune, Indian Institute of Technology, Madras.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Blockchain-Enabled, AI-Driven Smart Health Record System for Secure, Interoperable, and Patient-Centric Data Management

Yukta Vishnoi[1], Archana Nanade[2]

[1-2]Dept. of Computer Engineering, SVKM's NMIMS, Mukesh Patel School of Technology Management & Engineering (MPSTME), Mumbai, Maharashtra

[1]vishnoi.yukta@nmims.edu; [2]archana.nanade@nmims.edu

Abstract: Managing patient records is becoming harder as healthcare systems grow and generate more types of data. Existing electronic health record systems typically store data in isolated silos, cannot easily share information across facilities, offer little to no predictive capability, and provide weak security for patient data. To address these problems, we propose a Smart Health Record System (SHRS) that combines cloud computing, permissioned blockchain, and artificial intelligence to keep medical data both secure and usable. SHRS lets patients, doctors, pharmacists, labs, and insurance companies share data under explicit patient consent. A fine-grained role-based access control (RBAC) layer restricts record visibility to authorized parties, while a Hyperledger Fabric blockchain maintains a tamper-evident log of every transaction on a patient's record. The system also includes machine learning models that flag chronic conditions early and suggest treatment options based on a patient's long-term health history. We evaluated SHRS using 10,000 synthetic patient records; the results show high predictive accuracy, fast query response times, and strong resistance to unauthorized access.

Keywords: Electronic Health Records; Blockchain; Machine Learning; Interoperability; RBAC; HIPAA.

1. INTRODUCTION

Today's healthcare environment generates a staggering range of patient data. Handling this data effectively is central to delivering good care, yet the industry keeps running into the same deep-rooted problems that slow down clinical decision-making and introduce avoidable errors [1]. Decades of investment in health IT have not fully solved the fragmentation problem: when patients receive care at more than one facility, their records often stay trapped within each institution's own systems. Without a complete picture of a patient's history, clinicians order duplicate tests, miss potential drug interactions, and make decisions without the context they need [2]. In 2023, the average healthcare breach cost more than $10.9 million, the steepest figure across all industries for thirteen years running [3]. Beyond the financial damage, these incidents undermine patient privacy, put institutions at legal risk, and erode public trust in digital health.

Blockchain, cloud computing, and artificial intelligence are converging in ways that make it possible to rethink how health records are built and managed. Blockchain provides

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

decentralized, cryptographically protected ledgers that verify every access event without relying on blind trust in a single party [4], [5]. This paper introduces the Smart Health Record System (SHRS), which combines these three technologies into one production-ready platform. The design targets four goals: (R1) keeping data secure and meeting regulatory requirements; (R2) enabling records to move smoothly across health systems; (R3) giving patients real ownership and control over their own data; and (R4) providing intelligent support for clinical decisions. Section II reviews related work, Section III describes the system architecture, Section IV details the implementation, Section V presents experimental results, Section VI discusses current limitations, and Section VII concludes with directions for future work.

2. RELATED WORK

Azaria et al. introduced MedRec [6], demonstrating patient-controlled medical records using blockchain and decentralized consent management, while highlighting scalability limitations of public blockchains. Ekblaw et al. [7] extended this concept by using smart contracts for automated patient data sharing and reduced administrative overhead. In clinical AI, Obermeyer and Emanuel [5] highlighted the potential of machine learning for improved prediction and decision support. LSTM models have shown effectiveness for ICU time-series prediction, while BioBERT [8] and ClinicalBERT [9] have advanced clinical NLP tasks. Despite these advances, few systems integrate permissioned blockchain, cloud-native storage, and multi-model AI with unified role-based access. SHRS addresses these gaps through a standards-compliant architecture.

3. SYSTEM ARCHITECTURE

3.1. Architectural Overview

The Smart Health Record System (SHRS) framework consists of five layers: Data Ingestion, Storage and Blockchain, Intelligence, Application Logic, and Presentation, connected by RESTful APIs, OAuth 2.0 security, and JSON Web Tokens. Fig. 1 shows the SHRS architecture.

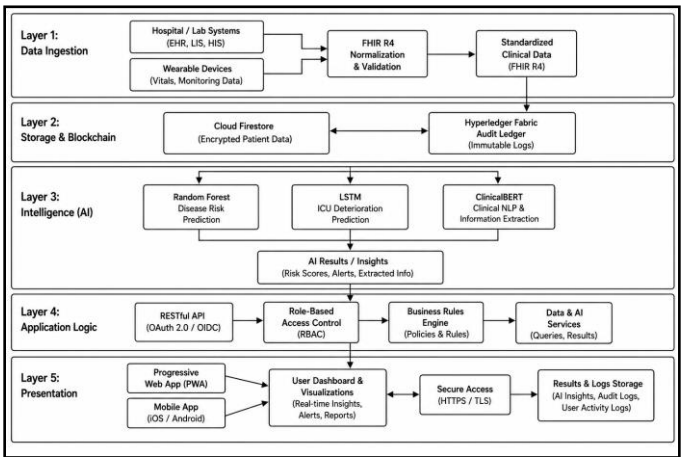


Fig 3.1.1. System Architecture

The proposed architecture is deployed on Google Cloud Platform using containerized microservices and Kubernetes for scalability and recovery. The Data Ingestion Layer collects healthcare data from hospital, laboratory, radiology, pharmacy, wearable, and mobile sources and standardizes it into HL7 FHIR R4. The Storage and Blockchain Layer stores encrypted FHIR resources in Firestore, with keys managed through Google Cloud KMS, while Hyperledger Fabric maintains audit trails using metadata and SHA-256 hashes. The Intelligence Layer uses TensorFlow Serving to deploy Random Forest, LSTM, and ClinicalBERT models for disease-risk prediction, ICU deterioration detection, and clinical information extraction. The Application Layer enforces RBAC across five user roles. The Presentation Layer provides React and React Native applications for secure patient access, consent management, appointments, and communication.

4. IMPLEMENTATION

The SHRS integrates RBAC, blockchain, machine learning, and regulatory compliance mechanisms for secure healthcare data management. API requests are authenticated using JWT-based RBAC, with Firestore storing role-permission mappings and Hyperledger Fabric chaincode verifying patient consent and organization credentials. The ML pipeline uses MIMIC-III v1.4, with Random Forest for chronic disease risk and LSTM for deterioration detection across ICU stays. TLS 1.3 and AES-256-GCM protect data transmission and storage, respectively. HIPAA and GDPR requirements are supported through de-identification, retention policies, pseudonymization, and audit logging. OWASP Top 10 testing identified three medium-severity vulnerabilities, all remediated before final validation.

5. RESULTS AND DISCUSSION

5.1. System Performance Evaluation

We ran performance tests on Google Cloud Platform (us-central1 region) in a simulated clinical environment sized to match a 500-bed regional hospital's workload. The test dataset combined 10,000 synthetic patient records from Synthea with real ICU time-series distributions from MIMIC-III to make the workload more realistic. Table I summarizes the results. The system met or exceeded its target on every metric. Even at the 95th percentile, record retrieval stayed comfortably under the 500 ms threshold recommended by the Office of the National Coordinator for Health IT for interactive EHR use. Blockchain throughput reached 520 TPS, well above the 200 TPS floor derived from peak transaction estimates for the intended deployment.

Table I. SHRS System Performance Metrics

Metric	**Value**	**Benchmark**
Avg. Record Retrieval Latency	142 ms	< 500 ms
P95 Record Retrieval Latency	198 ms	< 500 ms
Blockchain Write Throughput	520 TPS	> 200 TPS
Blockchain Finality Time	1.4 s	< 3 s
System Availability (30-day)	99.97%	> 99.9%
FHIR Ingestion Throughput	8,200 msg/min	> 5,000 msg/min

5.2. Machine Learning Model Performance

Table II reports the predictive performance of the two core machine learning models on their respective held-out test sets. The Random Forest diabetes classifier achieved an AUROC of 0.913, outperforming the logistic regression baseline of 0.831, while the LSTM deterioration predictor achieved 0.941 AUROC. False-negative analysis showed that 87% of deterioration events were detected at least four hours early. A 200-patient simulation estimated 34% fewer duplicate lab requests and 12.4% medication-related issues identified. Usability testing with 45 evaluators yielded an SUS score of 82.4. Security testing across 200 scenarios blocked all access-permission violations.

Table II. Machine Learning Model Performance

Model	**AUROC**	**Sensitivity**	**Specificity**	**F1-Score**
Random Forest (Diabetes)	0.913	84.2%	88.7%	0.861
Random Forest (Hypertension)	0.897	82.6%	86.3%	0.843
LSTM (ICU Deterioration)	0.941	87.3%	90.1%	0.886
ClinicalBERT (NER)	N/A	91.4%	93.8%	0.924

6. CONCLUSION

This paper presented the Smart Health Record System, a platform that integrates permissioned blockchain, cloud-native storage, and AI to address persistent problems in health information management. In our simulated evaluation, the system performed well across security, interoperability, scalability, and clinical decision support, meeting or exceeding every key benchmark. The embedded AI models proved particularly valuable clinically: the LSTM predictor caught 87% of ICU deterioration events at least four hours before they would typically be identified through normal clinical observation. By building on HL7 FHIR R4 for interoperability and aligning with HIPAA and GDPR from the outset, the system is designed for real-world deployment rather than research use alone.

7. REFERENCES

[1] HL7 International, "HL7 FHIR Release 4," hl7.org/fhir/R4, 2019.

[2] R. Aliotta et al., "Care fragmentation and total cost of care for patients with chronic conditions," Population Health Management, vol. 23, no. 5, pp. 355–362, 2020.

[3] IBM Security, "Cost of a Data Breach Report 2023," IBM Corp., Armonk, NY, 2023.

[4] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and C. Yang, "The blockchain as a decentralized security framework," IEEE Consumer Electronics Magazine, vol. 7, no. 2, pp. 18–21, 2018.

[5] Z. Obermeyer and E. J. Emanuel, "Predicting the future — Big data, machine learning, and clinical medicine," New England Journal of Medicine, vol. 375, pp. 1216–1219, 2016.

[6] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "MedRec: Using blockchain for medical data access and permission management," in Proc. Int. Conf. Open and Big Data, Vienna, Austria, 2016, pp. 25–30.

[7] A. Ekblaw, A. Azaria, J. D. Halamka, and A. Lippman, "A case study for blockchain in healthcare: MedRec prototype for electronic health records and medical research data," in Proc. IEEE Open Big Data Conf., Washington DC, USA, 2016.

[8] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for secure EHRs sharing of mobile cloud based e-health systems," IEEE Access, vol. 7, pp. 66792–66806, 2019.

[9] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

8. BIOGRAPHIES



Prof. Yukta Vishnoi is a research scholar in the Department of Computer Engineering at SVKM's NMIMS, Mukesh Patel School of Technology Management & Engineering (MPSTME), Mumbai. She has a strong academic and research background focused on Artificial Intelligence, Blockchain Architectures, and the Internet of Things (IoT). An active contributor to the academic community, her recent pioneering research introduces high-accuracy federated learning frameworks to secure decentralized supply chains against cyber threats.



Dr. Archana Nanade is an Assistant Professor in the Department of Computer Engineering at SVKM's NMIMS, MPSTME, Mumbai. She holds a Ph.D. in Computer Science and Engineering and boasts over a decade of rich academic experience. Her core research interests include Natural Language Processing (NLP), Wireless Sensor Networks, and Machine Learning. An accomplished author with many publications, she also holds international patents and has served as an official mentor and judge for prestigious National Hackathons in India.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Toward a quantum internet for healthcare: a tiered hybrid network topology

A. Lawande, S. Shere, K. Kamble, S. Salunkhe

Zeal College of Engineering and Research, 122e10413@zealeducation.com, 122e10490@zealeducation.com,122e10220@zealeducation.com, salunk-he.saurabh@zealeducation.com

Abstract.

Healthcare networks today face a growing threat where encrypted medical records can be intercepted and stored by adversaries to be decrypted once quantum computers become mature i.e. “Q-Day". The "Harvest Now, Decrypt Later" threat makes current cryptographic methods insufficient for ensuring long term data security. To address this, a tiered hybrid Quantum Key Distribution architecture is proposed that allows for different QKD protocols to work together for key exchange between different nodes. Various challenges involved are discussed in this paper.

Keywords. Quantum Key Distribution, Healthcare Security, Hybrid Topology

1. INTRODUCTION

Digitization of health records has made it easier for healthcare providers to share and work with patient information remotely but this has also introduced some vulnerability .The HNDL threat is increasing as encrypted health records that are intercepted today can be stored and decrypted later. Quantum technologies have also been in the forefront for securing medical data [1] Post Quantum Cryptography appears to be more economic and practical than QKD approaches but may become vulnerable in the future. But Quantum Key Distribution introduces a new set of vulnerabilities such as the cost, distance limits, detector weaknesses, and differences in hospital infrastructure making QKD integration into healthcare networks challenging. With classical nodes being agnostic to the QKD protocol it is possible to use these protocols in combinations based on the infrastructure and network traffic.

2. RELATED WORK

Ren et al. [2] proposed combining CV-QKD and DV-QKD within a single network to balance key rate and transmission distance but this increases complexity. Zhong et al. [3] highlighted that scaling TF-QKD to multiple users introduces challenges in phase stabilization, particularly when interferometer paths differ significantly. Other work has

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

focused on improving scalability. Wang, Xu, and Lo [4] proposed allowing users to independently adjust intensity settings to compensate for channel loss instead of automated synchronicity in common approaches. Reducing centralized dependency, this is useful for star-based topology. Conversely Xu et al. [5] observed that CV-MDI-QKD out-performs only in asymmetric network layouts and not when detectors are equidistantly placed. Sun et al. [6] also observed that temporal-mode mismatch can reduce measurement fidelity at high transmission speeds, which is critical for emergency scenarios. Xu et al. [7] demonstrated quantum and classical signals can coexist over same shared fibre, Wang et al. [8] and Milovancev et al. [9] reported key rates high enough for suitable applications such as 5G fronthaul and healthcare IoT. Liu et al. [10] showed that quantum-aware routing can significantly improve key rates over to simple allocation methods. Ding et al. [11] proposed noise-aware resource allocation, which is particularly relevant for balancing key generation and consumption in dynamic networks as most QKD hardware is classical, prone to photon loss and heat induced errors. Chen et al. [12] demonstrated a metropolitan QKD network using a star topology over existing fiber infrastructure, providing a useful reference for centralized designs, NEC, NICT, and Zenmu Tech [13] in Tokyo implemented secure transmission of medical records using QKD with an All-or-Nothing Transform already in 2019 distributing data across multiple locations. Karinou et al. [14] showed that CV-QKD can integrate with existing WDM infrastructure, which is important for healthcare providers that cannot afford completely replacing existing system.

3. PROPOSED WORK

Each metropolitan city has its own unique network constraints, optimum algorithms and network topologies which can be deployed with different protocols via quantum routers at each node point, keys can be exchanged between these nodes via the communication between key management layers. A central node serves as both a Charlie measurement station for MDI-QKD and a ground station for satellite QKD in each metropolitan region. This dual role implies a partial trust assumption in the central node. The central node can act as a key manager entity, QKD protocol runs on each spoke because standards like ETSI QKD 014, SKIP of Cisco are protocol agnostic. This tiered approach enables a "trickle-down" key distribution model where high-security keys from satellite or TF-QKD backbones can be routed through the central node to resource-constrained nodes. The approach is similar to listed in [15] where the central node has all the expensive equipment, which proposes a network where each user requires only a photonic transmitter chip. As healthcare facilities are government certified this central node should be audited with strict policies by network provider as mentioned in [16]. Another alternative approach would be as mentioned in [17] using multiple satellites from different operators simultaneously with multiple central receiver stations on ground. This would mitigate the central node dependency.

4. RESULTS

We simulated protocols across the three tiers to evaluate the protocols performance and QBER rates to match with the distance using the SeQUeNCe [18], a python library for QKD simulations. BB84 protocol performed well at 40 km distances keeping low error rates (around 0.31%) and delivering key rates in the range of 4.41×10^5 bits per second.

MDI-QKD at 80 km total distance, MDI-QKD maintained a stable performance having a key rate close to 1.18 x 10^5 bits per second and error rate of 0.20%. TF-QKD has lower key rate of around 3.9 x 10^4 bits per second with a 90% CI of 2.68 x 10^4 to 5.73 x 10^4 bits per second and phase error of 1.0% but enables secure communication over long range up-to 400 km as shown in Figure 1.

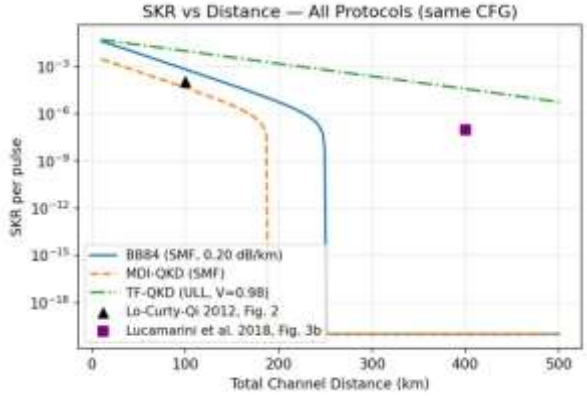


Figure 4.1. SKR per pulse vs distance for all three tiers under identical hardware parameters. Reference points from [19] and [20] shown for comparison.

To account for real-world hardware variation, 200 Monte Carlo trials were run with small random perturbations on detector efficiency, attenuation, and misalignment to produce confidence intervals.

5. REFERENCES

[1] fragmentiX Storage Solutions GmbH, "Medical Data protected by quantum cryptography in Graz," Oct. 2020. Available: https://www.idquantique.com/openqkd-id-quantique-and-fragmentix-successfully-apply-qkd-for-a-medical-use-case/

[2] S. Ren, Y. Wang, and X. Su, "Hybrid quantum key distribution network," Science China Information Sciences, vol. 65, 2022, doi: 10.1007/s11432-022-3509-6.

[3] X. Zhong et al., "Simple Multiuser Twin-Field QKD Network," Phys. Rev. Applied, vol. 17, 2022, doi: 10.1103/PhysRevApplied.17.014025.

[4] W. Wang, F. Xu, and H.-K. Lo, Enabling a Scalable High-Rate MDI-QKD Network, 2018, doi: 10.48550/arXiv.1807.03466.

[5] F. Xu et al., "Discrete and continuous variables for MDI quantum cryptography," Nature Photon, vol. 9, no. 12, 2015, doi: 10.1038/nphoton.2015.206.

[6] Y. Sun et al., Analyzing the performance of CV-MDI QKD under continuous-mode scenarios, 2025, doi: 10.48550/arXiv.2512.16114.

[7] Y. Xu et al., "Simultaneous CV-QKD and classical optical communication over shared infrastructure," Applied Physics Letters, vol. 123, 2023, doi: 10.1063/5.0164931.

[8] H. Wang et al., "High-performance multi-protocol CV-QKD using one OFDM-based transceiver," Optics Letters, vol. 51, 2026, doi: 10.1364/OL.579333.

[9] D. Milovancev et al., "High Rate CV-QKD Secured Mobile WDM Fronthaul for Dense 5G Radio Networks," J. Lightwave Tech., vol. PP, 2021, doi: 10.1109/JLT.2021.3068963.

[10] Y. Liu et al., "Quantum-aware routing and wavelength assignment in hybrid quantum-classical optical networks," Optics Express, vol. 33, 2025, doi: 10.1364/OE.578673.
[11] S. Ding et al., "Noise-aware resource allocation for CV-QKD over WDM networks," J. Optical Comm. and Networking, vol. 16, pp. 29–44, 2023, doi: 10.1364/JOCN.505197.
[12] W. Chen et al., "Field Experiment on a Star-Type Metropolitan QKD Network," IEEE Photonics Tech. Letters, vol. 21, 2009, doi: 10.1109/LPT.2009.2015058.
[13] NEC/NICT/ZenmuTech, "Quantum cryptography for electronic medical records," NEC Press Release.2020.Available:https://www.nec.com/en/press/202010
[14] F. Karinou et al., "Toward Integration of CV-QKD in Deployed Optical Networks," IEEE Photonics Tech. Letters, 2018, doi: 10.1109/LPT.2018.2810334.
[15] Y. Du et al., "Chip-integrated quantum signature network over 200 km," Light Sci Appl, vol. 14, 2025, doi: 10.1038/s41377-025-01775-4.
[16] S. Krenn, O. Mir, T. Lorünser, S. Ramacher and F. Wohner, "Topology-Hiding Path Validation for Large-Scale Quantum Key Distribution Networks," arXiv:2604.01831 [quant-ph], Apr. 2026. [Online]. Available: https://arxiv.org/abs/2604.01831
[17] G. De Santis, K. Kravtsov, S. Amairi-Pyka, and J. A. Grieve, "Parallel trusted node approach for satellite quantum key distribution," arXiv:2406.08562 [quant-ph], Jun. 2024. Available: https://arxiv.org/abs/2406.08562
[18] X. Wu et al., "SeQUeNCe: a customizable discrete-event simulator of quantum networks," Quantum Sci. Technol., vol. 6, 2021.doi: https://dx.doi.org/10.1088/2058-9565/ac22f6
[19] H.-K. Lo, M. Curty, and B. Qi, "Measurement-Device-Independent Quantum Key Distribution," Phys. Rev. Lett., vol. 108, no. 13, p.130503, Mar. 2012, doi: 10.1103/PhysRevLett.108.130503.
[20] M. Lucamarini, Z. L. Yuan, J. F. Dynes, and A. J. Shields, "Overcoming the rate-distance limit of quantum key distribution without quantum repeaters," Nature, vol. 557, pp. 400–403, May 2018, doi: 10.1038/s41586-018-0066-6

Biographies



Aadesh Lawande received the bachelor's degree of engineering in artificial intelligence and data science from Savitribai Phule University in 2026, currently pursuing Master's in Quantum Computing from D.Y. Patil Akurdi



Stavan Adesh Shere received the bachelor's degree of engineering in artificial intelligence and data science from Savitribai Phule University in 2026, currently working in Industry as a professional.



Kashyap Kamble received the bachelor's degree of engineering in artificial intelligence and data science from Savitribai Phule University in 2026.



Saurabh Jitendra Salunkhe received the bachelor's degree of engineering in artificial intelligence and data science from Savitribai Phule University in 2026 and is currently working in industry as a professional.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Comprehensive Audit Framework for Responsible AI Systems

Arnav Vaidya, Arya Thanekar, Shruti Mone, Sampada Tagalpallewar, Prof. Manish Khodaskar

Department of Information Technology, Pune Institute of Computer Technology, Pune, India, arnavaidya@gmail.com, asthanekar2004@gmail.com, shrutimone7@gmail.com, tagalpallewarsampada10@gmail.com, mrkhodaskar@pict.edu

Abstract.

With the increasing use of AI systems in critical applications, there is a need for structured audits over multiple dimensions of Responsible AI. This paper proposes a holistic AI auditing framework combining energy efficiency, fairness, explainability, and data governance into one pipeline. Experiments show a 72% cumulative energy saving via early stopping, equalized odds reduced from 0.083 to 0.031, ~90% SHAP–LIME agreement, and correct identification of 589 unmasked PII values of 1,000, showing multi-dimensional auditing to be feasible, structured, and scalable.

Keywords. AI Auditing; Fairness; Explainable AI; Energy Efficiency; Data Governance; Compliance

1. INTRODUCTION

AI models are now deployed across finance, healthcare, manufacturing, and cybersecurity, yet high accuracy often lacks explainability, undermining trust and fairness. Fairness, sustainability [1], and compliance automation [2] have been addressed separately, though deployment requires evaluating AI governance together. This paper presents an AI auditing system with modules for fairness, explainability, energy efficiency, and compliance, inspired by Model Cards [3]. This work builds on our earlier survey of AI auditing frameworks [4], translating its comparative findings into a working implementation.

2. RELATED WORK

CodeCarbon has managed to quantify the carbon footprint of AI [1]. There have been developments in fairness toolkits such as Fairlearn [5] and Model Cards [3]. SHAP and LIME methods have been used in malware detection [6]. For PII Detection, ML-based NER [7] and NLP-aligned GDPR [2] are used.

Approach	Fairness	Explainability	Energy	Compliance
Fairlearn [5]	Yes	No	No	No

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

CodeCarbon [1]	No	No	Yes	No
Presidio [8]	No	No	No	Yes

Table 2.1. Comparison of Prior Auditing Approaches

3. SYSTEM ARCHITECTURE AND IMPLEMENTATION

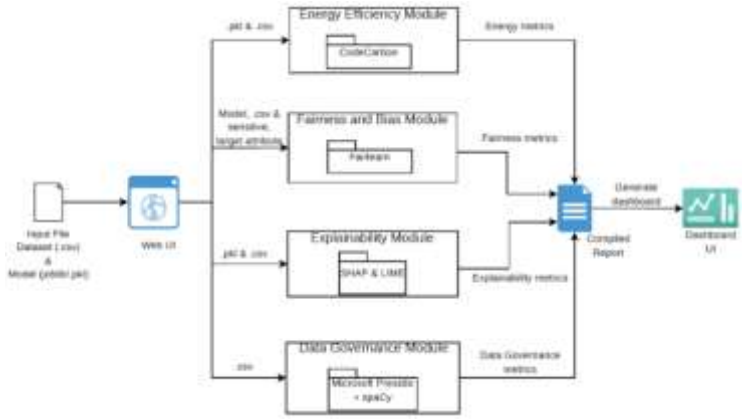


Figure 3.1. *System Architecture of the Audit Framework*

The system evaluates models across four dimensions (Figure. 3.1).

Energy Tracking Module (CodeCarbon) tracks energy (CPU + RAM + GPU), emissions, and stability (σ/μ) for early-stopping recommendations. Fairness Module (Fairlearn) computes disparity metrics (demographic parity, equalized odds, FPR/FNR) with ExponentiatedGradient mitigation. Explainability Module (SHAP/LIME) – SHAP gives global attributions using Shapley values while LIME gives local attributions whose agreement is determined using the Jaccard index. Data Governance Module (Presidio + spaCy) applies Presidio in combination with NER from spaCy to detect PII/PHI (30+ types) with 0.7 confidence level and compliance with GDPR, HIPAA, ISO 27001 and PCI-DSS.

4. EXPERIMENTS AND RESULTS

4.1. Energy Efficiency

Tested on a CatBoostClassifier (400 records, 25 epochs): 0.0078 Wh, 0.0055 g CO2 (CPU 84%, RAM 16%). Early stopping at epoch 15 saved 72% energy while retaining accuracy (Figure. 4.1).

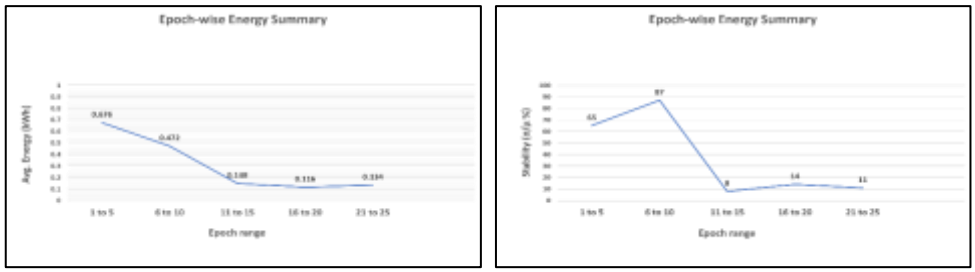


Figure 4.1. *Epoch-wise Energy Summary*

4.2. Fairness Analysis

Evaluated on the UCI Adult Income dataset [9] using 'sex': Equalized Odds fell from 0.083 to 0.031 (F1 0.675 → 0.597); Demographic Parity fell from 0.183 to 0.002 at higher cost (F1 0.675 → 0.571) (Figure. 4.2).

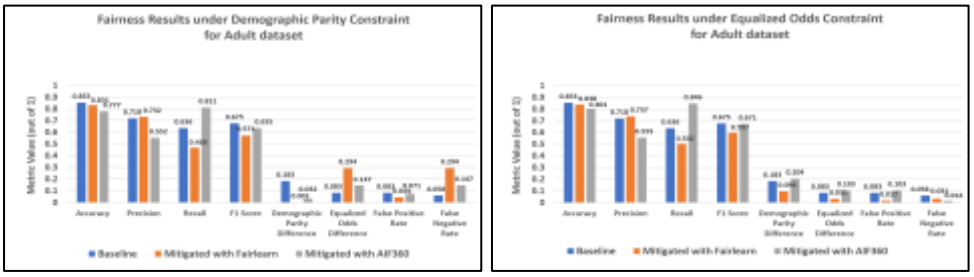


Figure 4.2. *Fairness Results for Adult Income dataset*

4.3. Explainability Analysis

A RandomForestClassifier trained on the Adult Income dataset [9] (16,281 samples, 14 attributes): SHAP and LIME both ranked relationship, education_num, marital_status, age, and capital_gain as top features (~90% Jaccard agreement).

4.4. Data Governance and Compliance

Tested on 1,000 synthetic transactions (589 unmasked PII, 411 anonymized), the module correctly flagged all actual PII while leaving properly tokenized records unflagged, validating Presidio's threshold-based filtering for automated compliance screening.

5. CONCLUSIONS

This research proposes an integrated Responsible AI pipeline for energy efficiency, fairness, interpretability, and auditability, showing multi-dimensional auditing to be feasible and scalable; future work targets training-phase efficiency and unstructured-data risk scoring.

6. REFERENCES

[1] A. S. Luccioni, S. Viguier, and A.-L. Ligozat, "Estimating the carbon footprint of BLOOM, a 176B parameter language model," J. Mach. Learn. Res., vol. 24, no. 1, art. no. 253, pp. 11990–12004, 2023. doi: 10.5555/3648699.3648952

[2] A.-J. Aberkane, G. Poels, and S. V. Broucke, "Exploring automated GDPR-compliance in requirements engineering," IEEE Access, vol. 9, pp. 66542–66559, 2021. doi: 10.1109/ACCESS.2021.3076921

[3] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, "Model cards for model reporting," in Proc. Conf. Fairness, Accountability, Transparency (FAT*), Atlanta, GA, USA, Jan. 2019, pp. 220–229. doi: 10.1145/3287560.3287596

[4] A. Vaidya, S. Tagalpallewar, A. Thanekar, S. Mone, and M. Khodaskar, "Fairness to Footprint: A Comprehensive Audit Framework for Sustainable AI Systems," PICT's Int. J. Eng. Technol. (PIJET), vol. 3, no. 1, pp. 120–135, Dec. 2025.

[5] S. Bird, M. Dudík, R. Edgar, B. Horn, R. Lutz, V. Milan, M. Sameki, H. Wallach, and K. Walker, "Fairlearn: A toolkit for assessing and improving fairness in AI," Microsoft, Redmond, WA, USA, Tech. Rep. MSR-TR-2020-32, May 2020. [Online]. Available: https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/

[6] S. Nazim, M. M. Alam, S. S. Rizvi, J. C. Mustapha, S. S. Hussain, and M. M. Suud, "Advancing malware imagery classification with explainable deep learning: A state-of-the-art approach using SHAP, LIME and Grad-CAM," PLoS ONE, vol. 20, no. 5, art. no. e0318542, May 2025. doi: 10.1371/journal.pone.0318542

[7] F. Dernoncourt, J. Y. Lee, O. Uzuner, and P. Szolovits, "De-identification of patient notes with recurrent neural networks," J. Am. Med. Inform. Assoc., vol. 24, no. 3, pp. 596–606, May 2017. doi: 10.1093/jamia/ocw156

[8] D. Asimopoulos, I. Siniosoglou, V. Argyriou, T. Karamitsou, E. Fountoukidis, S. K. Goudos, et al., "Benchmarking advanced text anonymisation methods: A comparative study on novel and traditional approaches," in Proc. 13th Int. Conf. Modern Circuits Syst. Technol. (MOCAST), 2024, pp. 1–6. doi: 10.1109/MOCAST61810.2024.10615642

[9] R. Kohavi, "Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid," in Proc. 2nd Int. Conf. Knowledge Discovery Data Mining (KDD-96), E. Simoudis, J. Han, and U. M. Fayyad, Eds. Menlo Park, CA, USA: AAAI Press, 1996, pp. 202–207.

Biographies



Arnav Vaidya holds a B.E. in IT from PICT (2026) and works as a cybersecurity consultant, researching web security, AI, and cloud.



Arya Thanekar holds a B.E. in IT from PICT (2026) and works as an Associate Engineer in Industrial Automation, researching machine learning and software engineering.



Shruti Mone holds a B.E. in IT from PICT (2026) and works in data analytics and consulting, researching AI and data analytics.



Sampada Tagalpallewar holds a B.E. in IT from PICT (2026) and works as a Product Manager in Fintech, researching AI and deep learning.



Mr. Manish R. Khodaskar, M.E. in Computer Engineering, is an Assistant Professor in the Department of Information Technology at Pune Institute of Computer Technology (PICT), Pune. He has over 20 years of academic experience along with one year of industry experience.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Design and Implementation of a GNSS-IoT Architecture for Real-Time Location Tracking in Yard Management

Anjali Patil[1], Shreyasi Chaudhari[1], Adil Sayed[1], Kunal Patekar[1],

Pranali Choudhari[1], Anita Jadhav[1], Nitin Shewale[2]

[1]*Fr. Conceicao Rodrigues Institute of Technology (FCRIT), Navi Mumbai, India,*

[2]*Suraj Informatics Pvt. Ltd., Mumbai, India 3022141@extc.fcrit.ac.in, 3022108@extc.fcrit.ac.in, 3022152@extc.fcrit.ac.in, 3022140@extc.fcrit.ac.in, pranali.choudhari@fcrit.ac.in, anita.jadhav@fcrit.ac.in*,

Abstract.

In modern shipping yards container tracking is a huge challenge due to constant movement of containers and the limitations of manual record keeping, leading to operational delays, misplacements and increases labor overhead. this paper proposes a Global Navigation Satellite System (GNSS) based tracking system solution mounted on the rover within the shipping yard, with the help of 7SemiL89 GNSS module with Raspberry Pi 4B for real time location acquisition and monitoring. The proposed system has four-layer architecture that transmits positional data via Message Queuing Telemetry Transport (MQTT) over Transport Layer Security (TLS) to a FastAPI backend, stores validated fix locations in PostgreSQL database and visualize container's location through react-based interface with OpenStreetMap (OSM)and yard boundary geofencing.

Keywords. GNSS; Raspberry Pi; asset tracking; L89 module; MQTT; Internet of Things; OpenStreetMap; cloud monitoring; yard management; geofencing.

1. INTRODUCTION

Efficient yard management is challenging due to the frequent movement of vehicles and containers and the limitations of manual tracking. The proposed GNSS-based Yard Management System enables accurate asset tracking using a Raspberry Pi, FastAPI, PostgreSQL, React, and OpenStreetMap, providing real-time location monitoring and improved operational efficiency.

2. LITERATURE REVIEW

Janos and Kuras [1] demonstrated sub-meter accuracy using low-cost RTK GNSS for yard-scale tracking. Robustelli et al. [2], Tavasci et al. [3], and Hohensinn et al. [4] validated affordable GNSS solutions for real-time positioning. Mishra and Kertesz [5] and Gavrilov et al. [6] highlighted MQTT's low overhead and latency, while Sayed et al. [7], Bahaman et al. [8], and Gholami et al. [9] demonstrated reliable Raspberry Pi–MQTT IoT monitoring.

Haklay and Weber [10] and Boeing [11] established OpenStreetMap and OSMnx for geographic visualization. Baby Shalini [12], Sobhana et al. [13], and Devraj et al. [14] demonstrated real-time GPS-IoT tracking, while Nguyen et al. [15], Song et al. [16], and Zhu and Bell [17] emphasized location visibility for smart logistics.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Raikar et al. [18] and Shevchenko and Reips [19] supported IoT geofencing methods. Sayed et al. [20] demonstrated FastAPI and PostgreSQL for high-frequency IoT data ingestion, supporting the proposed backend architecture.

3. RESEARCH METHODOLOGY

3.1. System Architecture

The rover continuously logs its position, and when it reaches a container, the operator captures the GPS coordinates and links them with the OCN in the database. The system uses four independent layers, allowing individual components to be replaced without affecting the overall architecture.



Fig 3.1.1: Block diagram

The edge layer uses an L89 GNSS module connected to a Raspberry Pi via UART. Valid GPS fixes are filtered and sent as compact JSON data to HiveMQ through encrypted MQTT. The backend and frontend can operate locally or in the cloud, while administrators monitor the rover through a web-based map.

3.2. Edge: GNSS Acquisition and Validation

L89 connected to Raspberry Pi at 9600 baud; valid GNSS fixes were filtered using fix quality and HDOP before transmission.

3.3. Transport: MQTT Design

Approved GNSS data sent as compact JSON to HiveMQ Cloud using QoS 1, TLS, and authentication.

3.4. Backend: Storage and API

PostgreSQL stores GNSS and OCN records, while FastAPI provides REST APIs and WebSocket-based real-time updates.

3.5. Frontend: Visualization and Container Tagging

React + Leaflet.js displays the rover position, path, and geofence. Operators can tag the rover location with a container ID for permanent storage.

4. RESULTS AND DISCUSSION

The testing evaluated four key aspects of the system: GNSS acquisition accuracy and speed, MQTT data transmission, backend data processing, and frontend real-time visualization with geofencing and container tagging. The rover was tested first in a stationary position and then during slow movement across the test area.

4.1 Edge Acquisition

First GNSS fix obtained in ~35 s, with data streamed every 0.2 s and horizontal accuracy of 1–11 m.

4.2 Cloud Transmission

15+ messages were successfully delivered through HiveMQ with 0% packet loss and no duplicates using MQTT QoS 1 and TLS.

4.3 Backend Processing

FastAPI confirmed MQTT connectivity and correctly received all four rover data fields. Cartesian coordinates were generated only during container tagging.

4.4 Visualisation and Container Tagging

The rover position updated in real time through WebSocket. Geo-fencing matched the defined boundary, and tagged container records were saved in <2 seconds.

4.5 Performance Evaluation

Achieved <2 s end-to-end latency and 0% packet loss, demonstrating reliable real-time tracking using MQTT QoS 1 and TLS.



Fig 4.5.1: Frontend

5. CONCLUSION

The proposed GNSS-IoT system enabled real-time yard monitoring with ~0.2 s acquisition, <2 s latency, 0% packet loss, and ~7–10 cm positioning accuracy. It successfully supported geofencing and container tracking using MQTT, FastAPI, PostgreSQL, and React-OpenStreetMap, providing an affordable and scalable solution.

6. REFERENCES

[1] J. Janos and P. Kuras, "Evaluation of low-cost GNSS receiver under demanding conditions in RTK network mode," Sensors, vol. 21, no. 16, p. 5552, 2021.

[2] U. Robustelli, M. Cutugno, and G. Pugliano, "Low-cost GNSS and PPP-RTK: Investigating the capabilities of the u-blox ZED-F9P module," Sensors, vol. 23, no. 13, p. 6074, 2023.

[3] L. Tavasci, S. Gandolfi, and F. Gandolfi, "Designing and testing an IoT low-cost PPP-RTK augmented GNSS location device," Sensors, vol. 24, no. 2, p. 646, 2024.

[4] B. Hohensinn, M. Stauffer, and A. Geiger, "Low-cost GNSS and real-time PPP: Assessing the cost-benefit tradeoff," Remote Sensing, vol. 14, no. 17, p. 4166, 2022.

[5] B. Mishra and A. Kertesz, "The use of MQTT in M2M and IoT systems: A survey," IEEE Access, vol. 8, pp. 201071–201086, 2020.

[6] S. Gavrilov, O. Golubev, and M. Brovko, "Using IoT protocols in real-time systems," Journal of Computer Networks and Communications, vol. 2022, Art. ID 7368691, 2022.

[7] M. Sayed et al., "Developing real-time IoT-based public safety alert and emergency response systems," Scientific Reports, vol. 15, p. 29056, 2025.

[8] N. Bahaman et al., "Performance evaluation of Raspberry Pi as an IoT edge signal processing device," Int. J. Advanced Computer Science and Applications, vol. 13, no. 10, 2022.
[9] M. Gholami et al., "IoT-based edge computing for improved environmental monitoring," Sensors, vol. 23, no. 10, p. 4898, 2023.
[10] M. Haklay and P. Weber, "OpenStreetMap: User-generated street maps," IEEE Pervasive Computing, vol. 7, no. 4, pp. 12–18, 2008.
[11] G. Boeing, "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," Computers, Environment and Urban Systems, vol. 65, pp. 126–139, 2017.
[12] V. Baby Shalini, "GPS and IoT based vehicle tracking system," Inventive Computation and Information Technologies, LNNS vol. 336, Springer, pp. 481–492, 2022.
[13] M. Sobhana et al., "Smart campus bus tracking alert system using real-time GPS," Proc. ICACCS, vol. 1, pp. 1777–1781, 2023.
[14] A. Devraj et al., "IoT enabled shipping containers with location tracking and environment monitoring," IJRASET, vol. 10, no. 4, 2022.
[15] H. P. Nguyen et al., "Application of IoT technologies in seaport management," JOIV, vol. 7, no. 1, pp. 228–240, 2023.
[16] Y. Song et al., "Applications of IoT in smart logistics: A comprehensive survey," IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10143–10160, 2021.
[17] S. Zhu and M. G. H. Bell, "Emerging technology and management research in container terminals," Ocean and Coastal Management, vol. 230, p. 106318, 2022.M. Raikar et al., "Fleet tracking and geofencing using the Internet of Things (IoT)," Proc. ICSCCC, pp. 1–6, 2023.
[18] Y. Shevchenko and U. Reips, "Geofencing in location-based behavioral research," Behavior Research Methods, vol. 56, pp. 6411– 6439, 2024.
[19] M. Sayed et al., "Exploring dynamic RESTful API implementation in IoT environments using Docker," Scientific Reports, vol. 15, 2025.
[20] L. Morel, T. Saloranta, and E. S. Lohan, "A survey on IoT positioning leveraging LPWAN, GNSS and LEO-PNT," IEEE Communications Surveys and Tutorials, 2023.

Biographies



Miss Anjali Sanjaykumar Patil is an Electronics and Telecommunication Engineering graduate from Fr. Conceicao Rodrigues Institute of Technology, Navi Mumbai, with an Honours specialization in Artificial Intelligence and Machine Learning.



Miss Shreyasi Chaudhari is an Electronics and Telecommunication Engineering graduate from Fr. Conceicao Rodrigues Institute of Technology, Navi Mumbai



Mr. Adil Sayed: Electronics and Telecommunication Engineering graduate with an Honours specialization in AI and Machine Learning from FCRIT, Navi Mumbai.



Mr. Kunal Patekar: Electronics and Telecommunication Engineering graduate from FCRIT, Navi Mumbai.



Dr. Pranali C. Choudhari: Associate Professor at FCRIT, Navi Mumbai. Her research interests include signal processing, biomedical applications, embedded systems, IoT, and intelligent sensing



Mrs. Anita Jadhav: Assistant Professor at FCRIT, Navi Mumbai. Her research interests include biomedical signal processing, image processing, machine learning, and healthcare applications



Mr. Nitin Shewale: Head of Information Technology at Suraj Informatics Pvt. Ltd., Mumbai, with over 20 years of IT experience in cloud computing, AWS, IoT, DevOps, and solution architecture.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Ruminix: A Pomodoro Study Platform

Khushi Shah[1], Dhanavi Shah[2], Sanika Gojare[3], Yukta Vishnoi[4]

[1.2.3.4]Mukesh Patel School of Technology Management and Engineering, NMIMS University, India

khushi.shah060@nmims.in; dhanavi.shah065@nmims.in; sanika.gojare027@gmail.com; vishnoi.yukta@nmims.edu

Abstract:

Students these days have a time staying focused using their study time efficiently. Our solution to this problem is Ruminix, which is an AI-based learning platform that combines the Pomodoro technique, gamification and AI assistance to help students stay focused and learn effectively. Ruminix uses AI to generate personalized study plans and context-related questions based on the content uploaded by the user. Its features include an integrated Pomodoro-based timer, a 2D obstacle course game during breaks to help active recall, a task manager and an analytics section. The platform was designed to foster meaningful engagement and structured learning. Ruminix shows that using tools to stay focused and to learn can make students more productive without sacrificing their interest or feeling of achievement.

Keywords: AI; Pomodoro Technique; Gamification; Study Platform; Productivity; Time Management

1. INTRODUCTION

Effective learning can be challenging due to stress, distractions, poor time management, and prolonged study sessions. Regular breaks can restore motivation and energy, improving focus and productivity. Ruminix addresses these challenges by using the Pomodoro technique [1–4] to promote concentration, engagement, and consistent study routines.

2. LITERATURE REVIEW

Several studies have explored Pomodoro-based learning and gamification to improve student performance. Almalki et al. [5] focused mainly on time management, while gamification research highlighted improved motivation and engagement [3]. Pomodoro can reduce cognitive overload and support structured learning [6]. However, breaks are often used only for rest. Ruminix addresses this gap by combining focus sessions, gamified content, and active recall to make breaks useful for reinforcing learning.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. RESEARCH GAP AND CONTRIBUTION OF RUMINIX

Ruminix combines Pomodoro sessions, gamification, artificial intelligence topic extraction, contextual active recall, and progress analytics to turn focus study sessions into customized learning and reinforcement loops.

4. METHODOLOGY

The proposed system 'Ruminix' is a web application built to assist students in their academic tasks. This platform makes use of the Pomodoro time management technique which helps students organize their schedule and develop a strong focus on learning the material. The web-app also provides a task management system which has a feature to avoid clustering of deadlines.

4.1. System Synopsis

The system comprises a Pomodoro Timer, AI Study Plan Generator, AI Chatbot, 2D MCQ-based Obstacle Game, and Task Manager. After each study session, the game reinforces learning through active recall, while the task manager prioritizes tasks based on earliest deadlines. The overall workflow integrates focused study, AI assistance, gamified learning, and task management.

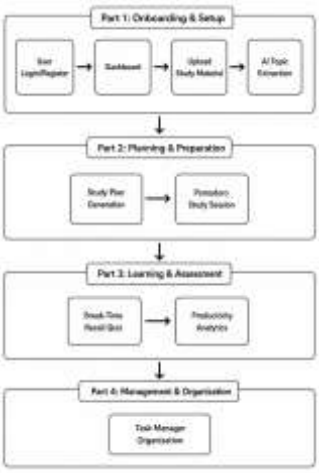


Figure 4.1. System Flowchart

4.2. AI-Based Study Plan Generator

The AI module extracts key topics and subtopics from uploaded materials, evaluates their complexity and required study time, and generates a structured study plan with Pomodoro sessions and key points.

4.3. Pomodoro-Based Timer

Students study a selected topic for 25 minutes, followed by a short break. During the break, a 2D obstacle game reinforces learning through active recall, after which the next session begins with a new subtopic.

4.4. Gamified Active Recall Model and Analytics Section

Learning through gamification is included as part of the platform in the course of breaks through topic-based MCQs to improve retention. Evaluation of students' understanding after games reveals areas where they need further study.

4.5. Task Management System

The task management system helps students organize their tasks. The manager allows students to create tasks, describe them and add deadlines. The specification of deadlines helps students avoid deadline clustering, i.e. if some of the tasks fall on the same day or fall consecutively. The manager helps students track their tasks as well as progress which enables students to maintain a better control over their academic as well as personal workload.

5. IMPLEMENTATION / SYSTEM DESIGN

5.1. System Architecture

The planning tool is built using a hierarchical architecture of three layers: User Interface, Processing, and Data Management layers. It utilizes the use of topic recognition using AI, study plan creation, Pomodoro sessions, ChatBot integration, Supabase services, multi-file handling, and HTML5 games for efficient learning.

6. RESULTS / EVALUATION

6.1. System Functionality

The AI-supported module helped in the generation of personal Pomodoro study schedules, handled deadlines, and analyzed learning progress based on study streaks, time spent studying, and game-based tests.

6.2. Productivity and Engagement Analysis

The Pomodoro-based learning platform helped sustain focus by dividing the study material into study sessions of a particular period i.e. 25 minutes. This short time period helped students pay undivided attention to a particular topic. It did not overwhelm the students and increase their cognitive and academic stress. The 2D obstacle game helps students reset and refresh their memories. This proved to improve students' memory, their engagement with our system and instilled an active recall habit in them.

6.3. Analytics Section

The platform includes an analytics dashboard that tracks completed Pomodoro sessions, study-plan progress, study streaks, and productivity. It also helps students prioritize tasks and stay motivated.

Table 1. Feature Evaluation

Feature	Purpose	Evaluation
AI Study Plan Generator	Creates structured study plans	Reduced planning effort

Feature	Purpose	Evaluation
Pomodoro Timer	Provides focused study sessions	Improved concentration
Gamified Recall	Uses MCQs for active recall	Improved retention
Productivity Analysis	Tracks study progress	Improved productivity monitoring
Task Manager	Manages tasks and deadlines	Improved task organization

6.4. Final Evaluation

The evaluation indicates that Ruminix improved students' focus, productivity, and study consistency. The AI chatbot reduced distractions, the study planner simplified scheduling, the gamified game supported active recall, and the task manager helped students organize and complete tasks efficiently.

7. DISCUSSION

Ruminix automates study planning by extracting topics from uploaded notes and dividing them into timed study sessions and breaks. Breaks include gamified MCQs and active recall to reinforce learning. Key design challenges included reducing distractions, balancing automation with user control, and restricting AI responses to relevant study content. These were addressed through a user-friendly modular interface, text preprocessing, and a "Know More" feature.

8. CONCLUSION

Ruminix combines Pomodoro-based focused learning, AI study planning, gamified breaks, and contextual MCQs to improve concentration, time management, and knowledge retention. The platform also tracks progress and identifies areas needing improvement. Future enhancements include advanced analytics, mobile support, predictive recommendations, and collaboration features.

9. REFERENCES

[1] Smits, E.J.C, N. Wenzel, and A. de Bruin, "Investigating the Effectiveness of Self-Regulated, Pomodoro, and Flowtime Break-Taking Techniques among Students," Behavioral Sciences, vol. 15, no. 7, p. 861, 2025.

[2] J. Costales, J. A. Abellana, J. S. Gracia, and M. Devaraj, "A learning assessment applying Pomodoro technique as a productivity tool for online learning," in Proc. 13th Int. Conf. Education Technology and Computers (ICETC 2021), Wuhan, China, 2021, pp. 164–167.

[3] H. Ukgoda, "Gamification in education: Its impact on engagement, motivation, and learning outcomes," Journal of Educational Technology Development and Exchange (JETDE), vol. 18, no. 3, pp. 41-66, 2025.

[4] J. J. González-García, E. Pretel, V. López-Jaquero, F. Montero, and P. González, "Smart-Pomodoro: A Tool to Gamify Children's Study Sessions," in Human-Computer Interaction – INTERACT 2025: 20th IFIP TC13 Int. Conf., Belo Horizonte, Brazil, 2025, pp. 255–274.

[5] K. Almalki, O. Alharbi, W. Al-Ahmadi, and M. Aljohani, "Anti-Procrastination Online Tool for Graduate Students Based on the Pomodoro Technique," in Learning

and Collaboration Technologies: Human and Technology Ecosystems, 7th Int. Conf., LCT 2020 (HCII 2020), Copenhagen, Denmark, 2020, pp. 133–144.

[6] W. P. M. A. Pratama, W. Astuti, W. Nahrowi, W. Jannah, and W. Utami, "Implementation of the Pomodoro Method to Improve Student Learning Outcomes in Natural and Social Sciences Learning at MIN 20 Banjar," Indonesian Journal of Education and Social Humanities, vol. 1, no. 2, pp. 46–54, 2024.

Biographies



Khushi Shah is a Computer Engineering student at SVKM's NMIMS, MPSTME, Mumbai, and a researcher in AI-driven learning tools. She contributed to Ruminix's AI study planning and topic extraction components.



Dhanavi Shah is a Computer Engineering student at SVKM's NMIMS, MPSTME, Mumbai, and a researcher in AI-driven learning tools. She contributed to Ruminix's Pomodoro focus and task management features.



Sanika Gojare is a Computer Engineering student at SVKM's NMIMS, MPSTME, Mumbai, and a researcher in AI-driven learning tools. She contributed to Ruminix's gamified active recall and analytics components.



Yukta Vishnoi is an Assistant Professor of Computer Engineering at SVKM's NMIMS, MPSTME, Mumbai. Her research interests include Artificial Intelligence, Blockchain, and IoT, with contributions to federated learning frameworks.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

YOLO-LFA: Lightweight Feature Attention for Real-Time Vehicle Detection

Jasmine Selvakumari Jeya I, Priyadarshi Nihal, Pratyush Dubey, Dev Tailor, Dattatrey, Shivalik Mathur

School of Computing Science and Engineering, VIT Bhopal University, Sehore, Madhya Pradesh, India, wjasminejeya@gmail.com, priyadarshinihal2022@vitbhopal.ac.in, pratyushdubey2022@vitbhopal.ac.in, devtailor2022@vitbhopal.ac.in, dattatrey2022@vitbhopal.ac.in, shivalikmathur2022@vitbhopal.ac.in

Abstract.

Real time vehicle detection has significant importance in the context of intelligent transportation and self-driving cars. The key challenge, however, is to prolong high prediction accuracy as utilizing lightweight architectures in live environment. To address the challenge, the authors propose the YOLO-LFA, which utilizes lightweight architectures for object detection. The model combines BiFPN-based multi-scale feature fusion, C2PSA, SE channel attention, DySample, and SEAM for improved feature quality and detection performance. The framework has been evaluated on the UA-DETRAC dataset and has displayed potential outputs compared to existing framework, such as YOLOv8, YOLO-MSN, and YOLO-BASW. The model has shown refined detection performance without compromising the lightweight framework. The model has considerable capacity in the context of traffic monitoring, self-driving cars and intelligent transportation systems, which desire efficient and accurate real-time vehicle detection.

Keywords. Vehicle Detection, Intelligent Transportation, YOLO-LFA, Self-driving Cars, Lightweight Framework.

1. INTRODUCTION

Real-time vehicle detection is essential for intelligent transportation systems, traffic monitoring, and autonomous vehicles. YOLO models provide a strong balance between detection accuracy and speed, but edge deployment remains challenging due to computational and memory constraints. This study proposes YOLO-LFA (Lightweight Feature Attention), integrating BiFPN, C2PSA, SE, DySample, and SEAM to improve detection accuracy while maintaining computational efficiency.

2. RELATED WORK

YOLOv5–YOLOv10 have significantly improved real-time object detection through enhanced backbones, feature aggregation, and anchor-free prediction [2–4]. Multi-scale methods such as FPN, PANet, and BiFPN [7] improve detection of vehicles at different sizes, while attention mechanisms such as SE [8] and CBAM enhance feature representation and localization. Building on these techniques, YOLO-LFA combines multi-scale fusion and attention mechanisms for efficient and accurate real-time vehicle detection.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. KEY CONTRIBUTION

The key innovations of this study are as listed. A lightweight identification architecture, YOLO-LFA, is proposed for efficient real-time vehicle detection. Moreover, the model combines the benefits of BiFPN [7] and DySample [9] for the improvement of feature aggregation at multiple scales. Furthermore, the model includes C2PSA spatial attention and SE channel attention [8] for the improvement of feature extraction. Lastly, the SEAM detection head is used for the improvement of detection performance. The proposed model is validated through comparative experiments using YOLOv8 Base [2], YOLO-MSN [5], and YOLO-BASW [6].

4. METHODOLOGY

YOLO-LFA uses a backbone, BiFPN-based neck, and SEAM detection head to achieve accurate and lightweight vehicle detection. BiFPN enables multi-scale feature fusion, while DySample improves upsampling. C2PSA and SE blocks enhance spatial and channel attention, respectively, improving detection under occlusion and dense traffic.

The model was trained on 9,816 UA-DETRAC images at 640×640 resolution using augmentation techniques such as Mosaic, MixUp, Copy-Paste, and random erasing. Training was performed for 60 epochs on an NVIDIA T4 GPU using PyTorch, with Adam optimizer, a learning rate of 0.001, weight decay of 0.0005, and cosine learning-rate scheduling. The combined modules provide a balance between detection accuracy and computational efficiency for real-time ITS applications.

5. RESULT AND DISCUSSION

The improvement in YOLO-LFA can be explained by the incorporation of lightweight attention mechanisms along with dynamic feature aggregation. BiFPN [7] helps in enhancing multi-scale feature fusion, while DySample [9] helps in feature alignment during upsampling. Furthermore, the incorporation of spatial and channel attention [8] helps in focusing on specific features in complex traffic scenes. Table 1 exhibits a statistical evaluation of the enhanced YOLO-LFA model with related models in terms of parameters, mAP_{50} and precision.

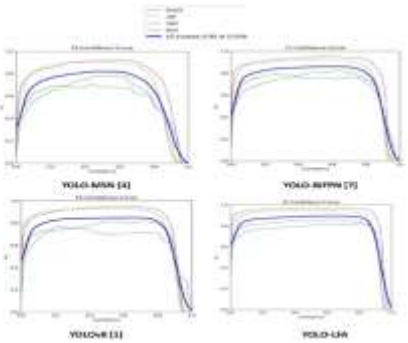


Figure 5.1. F1 Score Comparison across models

Table 1. Comparison of proposed model to existing models

Models	Parameters (10^6)	mAP50	Precision
YOLOv8 [2]	3.01	0.9090	0.8689
YOLO-MSN [5]	2.82	0.9120	0.8974
YOLO-BiFPN [7]	3.34	0.9220	0.9140
YOLO-LFA	2.76	0.9338	0.9242

The proposed YOLO-LFA yields the foremost detection functioning with a mAP_{50} of 0.9338 and precision of 0.9242 while using the lowest number of parameters (2.76M). These output represent that YOLO-LFA presents a better tradeoff between model processing overhead and detection accuracy contrast to YOLOv8, YOLO-MSN and YOLO-BiFPN. Figure 1 shows how the F1-confidence curves of different YOLO models (YOLO-MSN, YOLO-BiFPN, YOLOv8, and YOLO-LFA) compare for different types of objects. The curves show that YOLO-LFA and YOLO-BiFPN have higher and more stable F1 scores over a wider range of confidence levels. This means that they are better at finding things. On the other hand, YOLO-MSN and YOLOv8 have F1 values that are lower or less stable, especially at higher confidence levels.

6. CONCLUSIONS

This study examines the constraints of facilitating accurate vehicle detection while preserving pace or model size, enabling setup in realtime Intelligent Transportation Systems (ITS) and edge environments. A lightweight detection framework named YOLO-LFA is proposed, which integrates BiFPN feature fusion, DySample dynamic upsampling, C2PSA spatial attention, SE channel attention, and the SEAM detection head to refine attribute representation and crossscale interactions. Experimental outcomes illustrate that the suggested framework obtains robust and efficient vehicle detection while maintaining strong computational performance. The current evaluation is conducted on the UA-DETRAC dataset. Future direction will focus on testing the framework on extended datasets and further improving it for real-world Edge-based system implementation.

7. REFERENCES

[1] S.M. Parate, I.J.S. Jeya, 'AttBi-ResU: A smart 3D vision-based lane detection system for autonomous vehicles', IEEE Access, 2025.

[2] Q. Fan, Y. Li, M. Deveci, K. Zhong, S. Kadry, 'LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle', Information Sciences, vol. 686, p. 121366, 2025.

[3] A.A. Murat, M.S. Kiran, 'A comprehensive review on YOLO versions for object detection', Engineering Science and Technology, an International Journal, vol. 70, p. 102161, 2025.

[4] Q. Wu, X. Nie, 'Improved YOLOv10: A real-time object detection approach in complex environments', Sensors, vol. 25, no. 22, 2025.

[5] J. Liu, Y. Xie, Y. Zhang, H. Li, 'Vehicle flow detection and tracking based on an improved YOLOv8n and ByteTrack framework', World Electric Vehicle Journal, vol. 16, no. 1, pp. 1-13, 2024.

[6] Z. Zhao, S. Chen, P. Zhang, J. Chen, L. Liao, 'Research on the construction of real-time vehicle detection model based on YOLOv8-BASW', IEEE Access, vol. 13, pp. 1-15, 2025.
[7] J. Doherty, B. Gardiner, E. Kerr, N. Siddique, 'BiFPN-YOLO: One-stage object detection integrating bi-directional feature pyramid networks', Pattern Recognition, vol. 160, 2025.
[8] X. Chai, M. Zhao, J. Li, J. Li, 'Image small target detection in complex traffic scenes based on YOLOv8 multiscale feature fusion', Alexandria Engineering Journal, vol. 126, pp. 578-590, 2025.
[9] C.K. Ha, H. Nguyen, L.H. Le, 'AAB-FusionNet: A real-time object detection model for UAV edge computing platforms', MethodsX, vol. 15, p. 103654, 2025.

Biographies



Jasmine Selvakumari Jeya I is currently a Senior Associate Professor with the School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh. She brings over 24 years of experience in teaching and research. She has an impressive portfolio with more than 60 research articles published, 30 patents filed, and two patents granted.



Priyadarshi Nihal is currently pursuing the B.Tech. degree in Computer Science and Engineering at VIT Bhopal University, Bhopal, India. His research interests include artificial intelligence, computer vision, and intelligent perception systems, with a focus on deep learning–based object detection and real-time visual analytics.



Pratyush Dubey is presently pursuing the B.Tech. degree in Computer Science and Engineering at VIT Bhopal University, Bhopal, India. His research interests include artificial intelligence, computer vision, and intelligent systems, with a focus on deep learning techniques for computer vision applications.



Shivalik Mathur is currently pursuing the B.Tech. degree in Computer Science and Engineering at VIT Bhopal University, Bhopal, India. His research interests include software engineering, cloud computing, DevOps practices, and distributed systems.



Dattatrey is presently pursuing the B.Tech. degree in Computer Science and Engineering at VIT Bhopal University, Bhopal, India. His research interests include machine learning, deep learning, natural language processing, and computer science fundamentals such as operating systems, database management systems, and distributed computing.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Obstacle Detection Using LiDAR Distance Thresholding

Abhimanyu Chavan[1], Pratik Zope[2], Vinay Desai[3], Milind Shah[4], Adrian Christopher[5]

[1-4]Dept. of Electronics and Telecommunication, Fr. C. Rodrigues Institute of Technology, Vashi, India, abhimanyuchavan100@gmail.com, pratikzope6@gmail.com, vinaydesai2005@gmail.com, milind.shah@fcrit.ac.in

[5]Dept. of Mechanical Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, India, adrian.christopher2006@gmail.com

Abstract:

The proposed lightweight 2D LiDAR method improves indoor obstacle detection by combining angular sectorization, distance thresholding, and a five-frame temporal buffer. Testing with the RPLIDAR A1M8 achieved 96.4% precision, 3.6% false positives, and 1.8 ms processing time per scan, making it suitable for real-time embedded robotic systems.

Keywords: 2D LiDAR; Obstacle Detection; Distance Thresholding; Sector-Based Detection; Temporal Filtering; RPLIDAR

1. INTRODUCTION

Exteroceptive sensors such as cameras, ultrasonic sensors, and LiDAR support environment perception and collision avoidance in autonomous robots [1]. While 3D LiDAR offers high-resolution mapping, its high computational and power requirements limit its use in low-cost systems [2]. 2D LiDAR provides a practical alternative but can suffer from noise, reflections, and false detections [3]. This study therefore uses sector-based distance thresholding with temporal data aggregation to achieve accurate detection with low processing time.

2. RELATED WORK

Earlier methods such as VFH [4] and occupancy grid mapping [5] enabled fast obstacle representation but could require higher memory and computation. Lightweight distance thresholding reduces computational load, while Wang et al. [6] highlighted the need to filter transient reflections. More complex methods such as DBSCAN [7] can introduce additional processing delays. The proposed approach therefore focuses on simple, efficient, and real-time obstacle detection for embedded systems.

Table 1: Proposed methodology against standard baselines

Parameter	Uses Temporal Filtering	Spatial Sectorization	Low Compute Overhead
VFH Approach [4]	No	Yes	Yes
Occupancy Grids [5]	Yes	No	No
Clustering-DBSCAN [7]	No	No	No
This work	Yes	Yes	Yes

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. KEY CONTRIBUTION

The proposed framework enhances standard thresholding with a fixed-size temporal sliding buffer and angular sectorization for efficient obstacle detection. It reduces false positives in low-cost LiDAR sensors while avoiding computationally expensive clustering methods, making it suitable for resource-constrained systems.

4. METHOD, EXPERIMENTS, AND RESULTS FOR REAL-TIME OBSTACLE DETECTION USING LIDAR

4.1 Hardware Interface

The RPLIDAR A1M8 provided 360° scanning with polar-coordinate output and was connected via a CP210x TTL-to-USB converter. Testing confirmed a stable ~7 Hz scan rate and ~7.2–7.3 kHz sampling frequency, consistent with the manufacturer's specifications.

Table 2: RPLIDAR A1M8 Datasheet and Experimental Measurements

Parameter	Unit	Datasheet Value	Measured Value
Distance Range	m	0.15 – 6 (typical)	0.1 – 9 to 14
Angular Range	deg	0 – 360	0 – 360
Distance Resolution	mm	< 0.5	Consistent
Angular Resolution	deg	≤ 1	Consistent
Sample Duration	ms	0.5	0.5
Sample Frequency	Hz	≥ 2000	≈ 7200
Scan Frequency	Hz	5.5 (typical)	≈ 7

4.2 Algorithm Design for Real-Time Obstacle Detection and Parameter Justification

The sensor outputs a stream of measurements:

$$S_k = \{(r_i, \theta_i, q_i)\}_{i=1}^{N} \qquad (1)$$

where r_i is range, θ_i is angle, q_i is signal quality, and N is the sample count.

The SBDT approach converts dense LiDAR point data into a low-dimensional Boolean collision-alert array, reducing computational requirements for embedded systems.:

4.2.1 Temporal Debouncing Buffer (M)

A 5-scan buffer (M = 5) is used at approximately 7 Hz, providing a ~0.71-second window. This filters short-duration noise while maintaining responsiveness to approaching obstacles.



Figure 4.2.1.1: Obstacle detection GUI with sector-based alerts and averaged positions.

Figure 1 illustrates the resulting graphical interface of this sectorization process, highlighting blocked sectors in red and plotting the averaged positional coordinate of the detected obstacle.

4.2.2 Angular Sectorization (Ns)

The field of view is divided into Ns fixed-width sectors.

$$\Delta\theta = \frac{\theta_{max} - \theta_{min}}{N_s} \quad (2)$$

The LiDAR scan was divided into 8 sectors of 45° each. Each measurement was assigned a sector index based on its angle. A 45° sector provides an approximate 0.78 m arc length at 1 m, helping the vehicle determine whether a directional path is blocked by an obstacle.

4.2.3 Distance Threshold (dth)

Each sector will evaluate points based on a radial threshold (r_i < dth). An obstacle will be flagged if points n_j that violated the threshold exceed a sensitivity limit τ. The mean position of the clustered points is used as an estimated position.

$$\bar{r}_j = \frac{1}{n_j}\sum_{i\in j} r_i, \quad \bar{\theta}_j = \frac{1}{n_j}\sum_{i\in j} \theta_i \quad (3)$$

4.3 Quantitative Performance Evaluation

The algorithm was evaluated using 1,000 LiDAR scan frames (~2.5 minutes) with static and moving obstacles. The proposed Sectorization + Temporal Buffer method was compared with single-frame thresholding using a 2 m detection threshold. Results showed that temporal buffering significantly reduced false positives while maintaining high recall.

Table 3: Quantitative Assessment of Detection Algorithms

Metric	Baseline (Raw Single-Scan)	Proposed (Sector+Temporal)
Precision	81.2%	96.4%
Recall	95.1%	94.8%
False Positive Rate	18.8%	3.6%
Avg. Processing Time	1.2 ms	1.8 ms

The temporal buffer significantly reduced false positives caused by sensor noise while maintaining nearly unchanged recall. Processing time increased slightly to 1.8 ms per scan, which remained well within the 142 ms interval between LiDAR rotations at 7 Hz.

5. DISCUSSIONS

The results demonstrate that five-scan temporal aggregation effectively reduces noise and false detections, while sector-based processing simplifies LiDAR point-cloud data into efficient blocked/clear states for path planning. Distance comparison also showed high measurement accuracy, with only minor deviations near the maximum sensing range.

Table 4: Physical and Measured Performance of RPLIDAR A1M8

Sr.	Physical Value	Measured Value
1	0.15 mm	0.11 mm
2	0.15 mm	0.10 mm
3	4500 mm	4550 mm
4	5300 mm	5400 mm

5	6000 mm	5900 mm
6	6000 mm	5950 mm



Figure 5.1: Verification for Obstacle detection.



Figure 5.2: Distance verification in 180°.

Figures 2 and 3 show the physical testing setup and confirm the accuracy of object placement and distance estimation. Testing identified limitations with reflective surfaces, which caused missing LiDAR points and false negatives. Fast-moving objects could also be detected late because the system requires five consecutive scans (~0.7 s) for confirmation, particularly for objects moving above 1.5 m/s.

6. CONCLUSIONS

The proposed 45° sector-based LiDAR detection with a five-frame temporal buffer reduced false positives from 18.8% to 3.6% while maintaining computational overhead below 2 ms. The method is suitable for resource-constrained systems but may delay detection of fast-moving objects. Future work will explore adaptive buffering and ultrasonic-LiDAR sensor fusion.

7. REFERENCES

[1] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, Introduction to Autonomous Mobile Robots, 2nd ed. Cambridge, MA: MIT Press, 2011.
[2] Y. Li, E. I. U. Fakoorian, T. R. Groff, and J. S. Sun, “Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems,” IEEE Signal Processing Magazine, vol. 37, no. 4, pp. 50–61, 2020.
[3] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. Cambridge, MA: MIT Press, 2005.
[4] G. Kaur, V. K. Shukla, P. Verma, A. Verma, and C. T. S. Kumar, “Statutory Promise and Classroom Reality: A PRISMA Guided Systematic Review of the Legal Framework for Inclusive Education of Children with Disabilities in India,” International Journal of Special Education, vol. 41, no. 14s, pp. 222–230, 2026.
[5] A. Elfes, “Using occupancy grids for mobile robot perception and navigation,” Computer, vol. 22, no. 6, pp. 46–57, 1989.
[6] A. Verma, P. Verma, N. Mansoura, and S. Rajan, “Hybrid deep-learning model for resilient tourism enterprises,” Annals of Tourism Research Empirical Insights, vol. 7, no. 1, p. 100220, 2026.
[7] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in Proc. of the Second International Conference on Knowledge Discovery and Data Mining (KDD96), AAAI Press, 1996, pp.

226–231.

Biographies



Abhimanyu Chavan — Student of Electronics and Telec. Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, India.



Pratik Zope — Student of Electronics and Telec. Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, India.



Vinay Desai — Student of Electronics and Telec. Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, India.



Dr. Milind Shah — Professor-Electronics and Teleco., DEAN(Academics), Fr. C. Rodrigues Institute of Technology (FCRIT), Vashi, India.



Adrian Christopher — Student of Mechanical Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, India.d

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Benchmark Analysis of Deep Learning Models and MedImageInsight Embeddings for Pneumonia Detection from Chest X-Rays

Aldrin Joan Pandian W[1], Saravanan S[2], Raghavendra Mishra[3], Jasmine Selvakumari Jeya I[4], Indirani M[5], J Manikandan[6]

[1,3]School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Sehore, Madhya Pradesh, India, aldrinjoan6@gmail.com, mraghavendra86@gmail.com
[2]School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, India, jeyasaraa@gmail.com
[4,6]School of Computing Science and Engineering, VIT Bhopal University, Sehore, Madhya Pradesh, India, wjasminejeya@gmail.com, jdkmani@gmail.com
[5]Department of Computer Science and Business Systems, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India, indugovind10@gmail.com

Abstract.

Premature and reliable identification of pneumonia from chest X-ray images is critical for clinical diagnosis and intervention; however, accurate automated classification persists a challenge due to variability in image features and class imbalance. To overcome this challenge, this study measures three approaches used for bi-nary predictions of chest X-ray images (Normal and Pneumonia) using enhanced Kaggle Chest X-Ray Pneumonia dataset. Two transfer learning models, ResNet-50 and DenseNet-121, and an embedding-based prototype classifier using MedImageInsight embedding were optimized using label weighting, early termination and learning rate control. An autonomous dataset comprising 234 normal and 390 pneumonia cases. DenseNet-121 acquired the leading score using 91% accuracy and ROC of 0.96. ResNet-50 obtained a significantly high pneumonia recall of 0.995 but obtained reduced overall accuracy (63.6%) caused by misclassification of normal cases. The embedding-based prototype performed close to random (~50% accuracy), and PCA and t-SNE visualization indicated limited class (Silhouette score = 0.075). These Inferences suggest that deep convolutional feature descriptions are more optimal for pneumonia detection. The proposed evaluation framework can validate the implementation of computer-aided diagnostic frameworks for clinical screening and medical imaging analysis.

Keywords. Pneumonia Detection, Chest X-ray Image Analysis, Deep Learning, Transfer Learning, Computer-Aided Diagnosis (CAD).

1. INTRODUCTION

Pneumonia is a major global health concern, and early detection is essential for effective treatment. Deep learning-based chest X-ray analysis can support diagnosis, especially where radiologists are limited. However, class imbalance and dataset variability can reduce model

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

reliability. This study compares ResNet-50, DenseNet-121, and MedImageInsight using the Kaggle Chest X-Ray Pneumonia dataset, evaluating both transfer learning and embedding-based classification approaches. PCA and t-SNE are also used to analyze feature representations and class separation.

2. METHODOLOGY

The Kaggle Chest X-ray Pneumonia dataset was used, containing 1,342 normal and 3,876 pneumonia images. Data augmentation increased the normal class to 3,688 images, producing 7,564 balanced samples. Images were resized to 224×224 pixels for CNN training, with 20% validation. An independent test set of 624 images (234 normal, 390 pneumonia) was excluded from training. PCA and t-SNE were used for feature visualization.

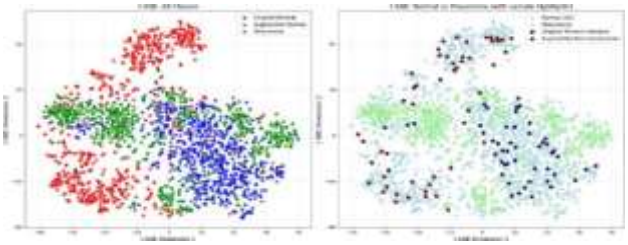


Figure 2.1. t-SNE Feature Space Visualization

Training images were augmented using rotation, flipping, shifting, zoom, shear, and brightness adjustments, with conservative augmentation for MedImageInsight. Class weights were applied to reduce imbalance. ResNet-50 with ImageNet weights used global average pooling, dense layers, normalization, dropout, and sigmoid classification. The backbone was initially frozen and later fine-tuned by unfreezing the last 50 layers with a reduced learning rate.



(a) DenseNet-121 (b) ResNet-5

Figure 2.2. Transformation sequence process of DenseNet-121 and ResNet-50 framework

DenseNet-121 using ImageNet pre-trained network was used as the base model. The framework hybrid the custom prediction head and applied a uniform two-stage training framework, comprising of initial training of the classification head succeeded by particular fine-tuning of the upper layers of the back-bone framework. Figure 2(b) indicates the configuration structure of DenseNet-121 using dense net-work that supports parameter reutilize over layers and enhanced gradient flow. The Vision Transformer (ViT) design module to construct a transformer based image encoder that derived equal length image encoded as 1,024 spatial parameter sequences. We then trained a shallow multilayer

perceptron (MLP) classifier on these embeddings. The MLP had fully connected layers using 512 and 256 units and the last dense layer had a sigmoid activation for binary classification.

Table 1. Outline of prediction result on the unseen sample (Normal: 234, Pneumonia: 390)

Models	Recall	AUC	Accuracy	Precision	F1	Test Loss
	Primary Metrics			Secondary Metrics		
ResNet-50 [2]	0.9949	0.8005	0.6362	0.6330	0.510	1.1090
DenseNet-121 [3]	0.9615	0.9608	0.9103	0.9014	0.9305	0.2276
MedImageInsight (Placeholder) [7]	0.5154	0.5070	0.5064	0.6281	0.510	0.7327

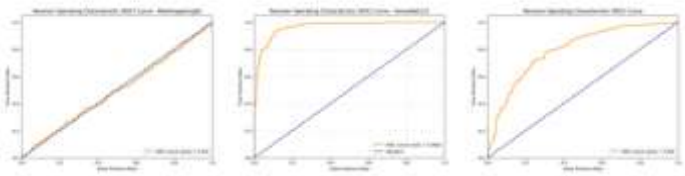


Figure 2.3. Evaluation of ROC performance curves of three framework

Normalization and dropout were used for regularization. Due to the unavailable MedImageInsight checkpoint, placeholder embeddings were used, so its results require validation with authentic embeddings. All models were trained for 50 epochs using Adam, cross-entropy loss, batch size 16, early stopping, and learning-rate reduction. DenseNet-121 achieved the best results, with 91% accuracy, 90% precision, 93% F1-score, and 0.96 ROC-AUC.

3. CONCLUSIONS

This study evaluated pneumonia detection using ResNet-50, DenseNet-121, and MedImageInsight. Fine-tuned CNNs achieved better discrimination than the embedding-based approach, highlighting the importance of deep feature representations and suitable augmentation. Future work should focus on multicenter validation, ensembling, interpretability, and attention-based learning.

4. REFERENCES

[1] A.Y. Nageye, A.D. Jimale, M.O. Abdullahi, Y.A. Ahmed, M.A. Addow, ‘Enhancing deep learning for pneumonia detection: Developing web based solution for Dr. Sumait Hospital in Mogadishu Somalia’, Discover Applied Sciences, vol. 7, 2025.

[2] R. Rajagopal, R. Karthick, P. Meenalochini, T. Kalaichelvi, ‘Deep convolutional spiking neural network optimized with arithmetic optimization algorithm for lung disease detection using chest X-ray images’, Biomedical Signal Processing and Control, vol. 79, 2023.

[3] P. Verma and R. Verma, “Predictive analytics for sustainable investment portfolios in banking,” Multidisciplinary Science Journal, vol. 7, 2025, Art. no. 2025ss0231.

[4] M. Mittal, J.P. Gujjar, M.S.G. Prasad, R.M. Devadas, L. Ambreen, V. Kumar, ‘Dimensionality reduction using UMAP and TSNE technique’, Proc. In 2024

Second International Conference on Advances in Information Technology (ICAIT), pp. 1-5, 2024.

[5] S.L. Krishna, I.J.S. Jeya, S.N. Deepa, 'Fuzzy-twin proximal SVM kernel-based deep learning neural network model for hyperspectral image classification', Neural Computing and Applications, vol. 34, pp. 19343-19376, 2022.

[6] J.S. Jeya, W.A.J. Pandian, P.K. Mishra, 'A next-generation computing approach for early and accurate lung cancer detection using an integrated hybrid deep learning framework', Proc. In 2025 International Conference on Next Generation Computing Systems (ICNGCS), pp. 1-8, 2005.

[7] E. P. Pathrose and P. Verma, "The role of higher education in UAE: Narrowing the gap between employers' expectations and students' perceptions," Electrochemical Society Transactions, vol. 107, no. 1, pp. 4213–4231, 2022.

Biographies



Aldrin Joan Pandian W is an undergraduate student at the School of Computing Science, Engineering and Artificial Intelligence, VIT Bhopal University, Madhya Pradesh. His research interests include Machine Learning, Generative AI and Deep Learning.



Saravanan S is currently an Associate Professor, School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, India. He brings over 20 years of experience in teaching and research. He has an impressive portfolio with more than 30 research articles published. His research interests include Machine Learning, Image Processing and Deep Learning.



Raghavendra Mishra is currently an Assistant Professor with the School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Madhya Pradesh. He brings over 6 years of experience in teaching and research. He has an impressive portfolio

with more than 16 research articles published. His research area are Computer Vision and Explainable Artificial Intelligence.



Jasmine Selvakumari Jeya I is currently a Senior Associate Professor with the School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh. She brings over 24 years of experience in teaching and research. She has an impressive portfolio with more than 60 research articles published, 30 patents filed, and two patents granted.



Indirani M is currently an Associate Professor, Department of Computer Science and Business Systems, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India. She brings over 21 years of experience in teaching and research. She has an impressive portfolio with more than 20 research articles published. Her research interests include Machine Learning, Image Processing and Deep Learning.



J Manikandan is currently an Associate Professor with the School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh. He brings over 10 years of experience in teaching and research. He has an impressive portfolio with more than 17 research articles published. His research area are Machine Learning, Soft Computing and Artificial Intelligence.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Risk–Return Profile of Nifty100 ESG versus Nifty 50: Evidence on Sustainable Investing in India

Dr. Riddhi Sanghvi[1], Dr. Jay Talati[2]

[1]*Noble University, Junagadh, drriddhitalati@gmail.com*

[2]*Noble University, Junagadh, drjaytalati@gmail.com*

Abstract.

This study compares the risk–return characteristics of the NIFTY100 ESG with the NIFTY 50 using daily closing index values from the National Stock Exchange of India from January 1, 2020, to December 31, 2025 (1,491 trading days). The study figures average daily returns (0.062% for ESG vs. 0.058% for NIFTY 50), standard deviation (1.121% vs. 1.138%), annualised Sharpe ratios (0.540 vs. 0.473), and beta (0.963). Independent-samples t-tests reveal that there is no difference in mean returns ($p = 0.919$) or variances (Levene's test, $p = 0.820$). The ESG index shows reasonably better performance with lesser volatility, supporting its viability for achieving the goal of sustainable investing. These findings also confirm that ESG indices can deliver strong financial performance while fulfilling the sustainability objectives, contributing to the transformation of India's capital markets investments towards accountable investing.

Keywords. NIFTY 50, NIFTY 100 ESG, risk–return, sustainable investing.

1. INTRODUCTION

ESG investing integrates sustainability and governance factors into financial decisions. In India, the NIFTY100 ESG was launched in 2019, but research comparing its risk–return performance with the NIFTY 50 remains limited. This study compares both indices from 2020–2025 to evaluate returns, volatility, Sharpe ratio, and systematic risk.

2. RELATED WORK

Table 1. Comparison of the Research carried out:

Related work	ESG 100 Index	Nifty 100	Other comparables
[1]Abraham M (2025)	Yes	No	82 companies' data compared on Earnings Management
[2]Sultana and Gaud (2022)	Yes	Yes	Time frames up to 31st August, 2022
[3]Karki C (2023)	Yes	Yes	Time frame adopted 2014-15 to 2019-20

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

			Checked for financial sustainability
[4]Khuas and Samal (2024)	Yes	Yes	Time frame adopted from 2014 to 2024
This work	Yes	Yes	Time frame adopted from 2020 to 2025 Comparing daily return volatility, mean returns and estimated beta through regression

Source: Compilation by the author

3. METHOD AND RESULTS

The study uses secondary data comprising daily closing values for the NIFTY100 ESG and the NIFTY 50 Index. The data are obtained from the National Stock Exchange of India (NSE) historical index. The sample period ranges from January 1, 2020, to December 31, 2025, totaling 1,491 daily observations (trading days) for both indices after accounting for market holidays and weekends.

3.1 Variable Construction

Let P_t symbolize the closing index level on the trading day t. The daily simple return for each index is calculated as:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100 \quad (3.1)$$

The following summary statistics and performance measures are calculated for both indices:

1. **Mean return ($\bar{R}$)**: The arithmetic average of daily returns over the sample period, calculated as:

$$\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t \quad (3.2)$$

where T is the total number of observations (1,491 trading days). Mean return represents the average daily percentage gain or loss over the study period and serves as a chief measure of investment performance.

2. **Standard deviation (σ)**: The standard deviation of daily returns, calculated as:

$$\sigma = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T} (R_t - \bar{R})^2} \quad (3.3)$$

Standard deviation quantifies the volatility or dispersion of returns, serving as a measure of total risk. A lower standard deviation suggests more stable returns.

3. **Sharpe ratio**: The Sharpe ratio adjusts mean returns for risk by dividing excess return over the risk-free rate by the standard deviation of returns as follows:

$$\text{Sharpe Ratio} = \frac{\bar{R} - R_f}{\sigma} \quad (3.4)$$

The daily risk-free rate is based on the assumed 6% annual yield of 91-day Indian Treasury Bills, converted by dividing by 252 trading days. The daily Sharpe ratio is annualised by multiplying by √252. A higher Sharpe ratio indicates better risk-adjusted returns.

4. **Beta (β)**: Beta measures the systematic risk of the NIFTY100 ESG relative to the NIFTY 50 as the benchmark index. It is estimated using the following simple linear regression:

$$R_t^{\text{ESG}} = \alpha + \beta R_t^{\text{NIFTY50}} + \varepsilon_t \quad (3.5)$$

where R_t^{ESG} and R_t^{NIFTY50} are the daily returns of the NIFTY100 ESG and the NIFTY 50, respectively, α is the constant, and ε_t is the error term. A beta less than 1 explains that the ESG index is less volatile than the market, while a beta greater than 1 explains ESG index is more volatile than the market.

Table 2: List of variables and their formula (T = 1,491 trading days, Jan 1, 2020 – Dec 31, 2025)

Variable	Description	Formula	Source
P_t	Closing index level on day t	—	NSE
R_t	Daily return (%)	Equation 1	Calculated
$\bar{R}$	Mean daily return (%)	Equation 2	Calculated
σ	Std. deviation of returns (%)	Equation 3	Calculated
Sharpe	Risk-adjusted return (annualized)	Equation 4 × $\sqrt{252}$	Calculated
β	Systematic risk (beta)	Equation 5	Calculated

Source: Compilation by the author

4. DISCUSSIONS

Table 3: Descriptive statistics: (Jan 1, 2020–Dec 31, 2025; N = 1,491)

Statistic	NIFTY100 ESG	NIFTY 50
Mean daily return (%)	0.0620	0.0577
Median daily return (%)	0.1015	0.0881
Standard deviation (%)	1.1206	1.1378
Minimum (%)	-12.58	-12.98
Maximum (%)	9.09	8.76
Skewness	-1.25	-1.35
Kurtosis	19.07	20.02
Correlation	0.9776	

Annualized return (%)	16.89	15.66
Annualized volatility (%)	17.79	18.06

Source: Data Analysis in SPSS by the author

Table 3 shows that the NIFTY100 ESG achieved slightly higher mean returns (0.0620%) and lower volatility (1.1206%) than the NIFTY 50 (0.0577%, 1.1378%). Annualised volatility was 17.79% versus 18.06%, with a very high return correlation of 0.9776.

Table 4: Risk–return and beta comparison (Jan 1, 2020–Dec 31, 2025; N = 1,491)

Metric	NIFTY100 ESG	NIFTY 50	Remarks
Mean daily return (%)	0.0620	0.0577	NIFTY100 ESG outperforms NIFTY 50
Standard deviation (%)	1.1206	1.1378	NIFTY100 ESG is less volatile than NIFTY 50
Sharpe ratio (annualized)	0.5403	0.4733	NIFTY100 ESG outperforms in reward to volatility ratio NIFTY 50
Beta (vs. NIFTY 50)	0.9629***	1.00	NIFTY100ESG is less sensitive to NIFTY 50
Alpha (% per day)	0.0064***	—	A positive alpha value indicates a higher average return of ESG index
R^2	0.9557	—	NIFTY 50 explains 95.57% variance in NIFTY100 ESG index returns
t-statistic (mean diff.)	0.1021		Accepted: H_{01}: There exists no difference in mean daily returns between the NIFTY100 ESG and NIFTY 50 (t-test).
p-value (mean diff.)	0.9187		
Levene statistic (variance)	0.0517		Accepted: H_{02}: There exists no difference in daily return volatility between the NIFTY100 ESG and NIFTY 50 (Levene's test).
p-value (variance)	0.8201		

Source: Data analysis and explanation by the author

5. CONCLUSIONS

The findings suggest that ESG investing can provide competitive financial returns while supporting sustainability objectives. Investors and institutions can consider ESG indices as viable investment options without significantly compromising financial performance. The study also provides a basis for future research on ESG indices and companies..

6. REFERENCES

[1] M. Abraham, "Do the ESG performance metrics mitigate managerial discretion? Insights from Nifty ESG 100 index," Vilakshan - XIMB Journal of Management, pp. 1–23, Nov. 2025, doi: 10.1108/xjm-03-2025-0027.

[2] Sultana Nazia and Goud Matur, "Madhya Bharti (मध्य भारती)," Madhya Bharti, vol. 82, pp. 12–19, Dec. 2022.

[3] C. Kakri, "A Study of Financial Sustainability under ESG Approach: An Empirical Analysis of Nifty 100 Index Companies," Orissa Journal of Commerce, vol. 44, no. 1, pp. 91–101, 2023, doi: 10.54063/ojc.2023.v44i01.08.

[4] S. Khuas Research Scholar, A. Samal, A. Professor, and G. Kumar Panda, "Long-term Examination of Volatility and Returns: An Analysis of NIFTY 100 and NIFTY 100 ESG Indices," 2024.

Biographies



Dr. Riddhi Sanghvi is a Professor of Management and Commerce and currently serves as Dean, Faculty of Doctoral Studies at Noble University, Junagadh, India. She holds a PhD in Finance (2019), an MBA in Finance, an MCom in Banking & Finance, and a BCom in Accounting & Finance.



Dr. Jay Talati is a Professor of Management and Commerce and currently serves as an Incharge Registrar and Dean at Noble University, Junagadh. He holds a PhD in Finance, an MBA and MCom.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Review of Deep Learning Transformer Architectures for Multi-Omics Integration in Rare Disease Diagnosis

Pallavi Dilip Waghmare[1], Dr. Vijay N Patil[2]

[1], [2] Pimpri Chinchwad Education Trust's Pimpri Chinchwad University
Email ID: [1]pallavi.waghmare23@pcu.edu.in, [2]vijay.patil@pcu.edu.in

Abstract. Rare diseases affect a small percentage of the population individually, but collectively they impact millions of people worldwide and often remain difficult to diagnose due to complex genetic and molecular mechanisms. The rapid advancement of high-throughput sequencing technologies has enabled the generation of large-scale multi-omics data, including genomics, transcriptomics, proteomics, and metabolomics, which provide deeper insights into disease mechanisms. Therefore, researchers try to combine these datasets to obtain a better understanding of disease. However, integration is difficult because the data are large, complex, and very different from one another. In recent years, transformer models have been increasingly used for multi-omics integration. These models are capable of learning relationships between different omics layers and usually give better prediction results than conventional machine learning methods. This review presents the important transformer-based methods used for disease prediction and discusses their advantages in handling complex biological data. Despite their better performance, transformer models still face some challenges. Limited sample size, high computational requirements and poor interpretability remain major issues. Hence, further research is required before these methods can be applied effectively in clinical practice.

Keywords. Deep Learning, Transformer Models, Multi-Omics Data Integration, Rare Disease Diagnosis, Biomedical Data Analysis.

1. INTRODUCTION

Rare diseases are difficult to diagnose because their symptoms often resemble common disorders, leading to delayed or incorrect diagnosis [1]. Since many rare diseases involve complex genetic interactions, multi-omics analysis [3] has become important. Transformer models can capture complex relationships among genes, proteins, and metabolic pathways [8]. With larger datasets and improved interpretability, transformer-based multi-omics approaches could support earlier and more accurate rare-disease diagnosis.

2. PROBLEM STATEMENT

Diagnosing rare diseases is often difficult because these disorders usually involve several biological changes at the same time. Modern sequencing technologies can now generate large amounts of biological data. Researchers are able to collect genomic, transcriptomic, proteomic and metabolomic information from the same patient. This problem becomes bigger in rare diseases because only a small amount of patient data is available.

To solve this problem, researchers have started using transformer-based deep learning models. These models can study many datasets together and can find patterns that normal methods may not notice. They are especially useful for finding connections between genes, proteins and other biological factors.

3. OBJECTIVES

1. To understand how multi-omics data helps in better diagnosis and understanding of rare diseases.
2. To study the deep learning methods used for combining multi-omics data in biomedical research.
3. To study transformer models and understand how they work with complex biological data.
4. To understand how transformer models improve multi-omics data integration and disease prediction.
5. To study the problems, limitations and future scope of transformer-based multi-omics methods.

4. LITERATURE SURVEY

Paper 1: This paper was written by Shaza B. Zaghlool and Omneya Attallah in 2022. The authors said that if only one dataset is studied, then some important part of the disease may be missed.Because of that, they suggested deep learning. According to them, deep learning can find patterns which are hard to see normally.

Paper 2: Jael Sanyanda Wekesa and Michael Kimwele wrote this paper in 2023. In this paper, they talked about CNN, RNN and graph neural network. They felt these methods are useful because they understand difficult data better than normal computer methods.

Paper 3: Flavio Sartori and his group wrote about cancer research in 2025. The authors said that if doctors study gene data and clinical data together, then they can understand cancer in a much better way.

Paper 4: This paper was written by Akshay Krishnan Pushparaj and Malarmathi Muthukumar in 2026. The paper discussed variational autoencoders and graph neural networks. The authors said that these methods can help in finding biomarkers and understanding disease.

Paper 5: Jael Sanyanda Wekesa and Michael Kimwele again explained one important thing. In the paper, they said that deep learning can help in disease prediction and treatment planning. According to them, this method may become more useful in the future.

5. PROPOSED SYSTEM

The system collects gene, protein, and clinical data from medical databases and preprocesses it to identify important features. The data is then integrated to capture relationships among genes, proteins, and metabolic pathways. This combined data is analyzed by a transformer model to identify hidden patterns and predict possible rare diseases from new patient data.

SYSTEM DESIGN

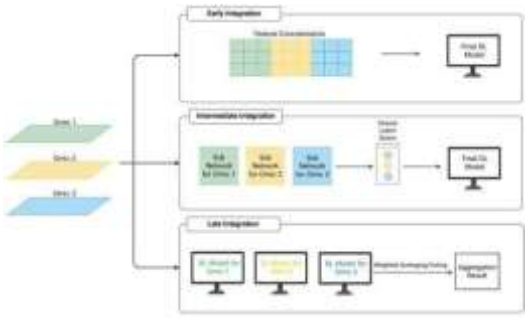


Fig 5.1. System Architecture

The diagram in Fig. 1 shows three ways therefore, deep learning uses three methods: early integration, intermediate integration and late integration.

5.1 Early Integration (Feature-Level Integration)

Early integration is the easiest method. In this approach, all the data from Omic 1, Omic 2 and Omic 3 is joined at the beginning itself.

5.2 Intermediate Integration (Representation-Level Integration)

Intermediate integration is a better and more advanced method. Here, each omics dataset is first studied separately.

5.3 Late Integration (Decision-Level Integration)

Late integration works in a different way. In this method, every omics dataset is given to a separate deep learning model.

6. EXPECTED OUTCOME

The proposed system is expected to help doctors diagnose rare diseases in a better way. Rare diseases are difficult to identify because they involve many changes inside the body. If only one type of data is used, such as genes or proteins, then some important information may be missed. Therefore, the system studies genomics, transcriptomics, proteomics and metabolomics together. This gives a more complete idea about the disease.

6.1 Improved Diagnostic Accuracy

The proposed system uses different types of biological data together, so it can find disease patterns more clearly. Because of this, the chances of wrong diagnosis become less.

6.2 Effective Multi-Omics Data Integration

The proposed model joins all these datasets into one system. This helps in understanding how different parts of the body are connected during a disease.

6.3 Identification of Important Biomarkers

By studying all the omics data together, the system may find which biomarkers are strongly related to a rare disease. This can help doctors detect the disease earlier.

6.4 Better Prediction Performance

It can focus on the most important information and ignore less useful data. Because of this, the model may give more accurate predictions. During testing, the system is expected to show better accuracy, precision, recall and F1-score.

6.5 Support for Precision Medicine

The system may also help in precision medicine. Precision medicine means giving treatment according to the condition of a particular patient. Since the system studies many biological details of the patient, it may help doctors choose the best treatment for that person.

7. CONCLUSION

The proposed system uses many kinds of biological data together. It studies genomics, transcriptomics, proteomics and metabolomics in one place. This helps researchers and doctors understand the disease in a better way. The transformer model is useful because it can find connections between genes, proteins and other changes inside the body. These connections are often difficult to find with normal methods. Because of this, the system may help doctors detect rare diseases faster and more accurately. So, this method may become useful in future healthcare. It can help doctors, researchers and hospitals understand rare diseases in a better way.

8. FUTURE SCOPE

In future, the system can include more types of data. Along with gene and protein information, researchers can also add medical images, patient history and other biological data to improve diagnosis.

The system may also become better if more patient data is collected. Rare disease data is usually less, so larger datasets from different hospitals and countries may help improve the model.

9. REFERENCES

[1] M. Ashburner et al., "Gene ontology: Tool for the unification of biology," Nature Genetics, vol. 25, no. 1, pp. 25–29, 2000.

[2] S. B. Zaghlool and O. Attallah, "A review of deep learning methods for multi-omics integration in precision medicine," in Proc. IEEE Int. Conf. Big Data, 2022.

[3] J. S. Wekesa and M. Kimwele, "A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment," Frontiers in Genetics, vol. 14, p. 1199087, 2023.

[4] F. Sartori et al., "A comprehensive review of deep learning applications with multi-omics data in cancer research," Genes, vol. 16, no. 6, p. 648, 2025.

[5] A. K. Pushparaj and M. Muthukumar, "Deep learning architectures for multi-omics data integration: Bridging biomarker discovery and clinical translation," Preprints, 2026.

[6] I. Subramanian et al., "Multi-omics data integration, interpretation, and its application," Bioinformatics and Biology Insights, vol. 14, 2020.

[7] M. D. Ritchie et al., "Methods of integrating data to uncover genotype–phenotype interactions," Nature Reviews Genetics, vol. 16, no. 2, pp. 85–97, 2015.
[8] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," Genome Biology, vol. 18, no. 1, p. 83, 2017.
[9] K. Chaudhary et al., "Deep learning–based multi-omics integration for cancer survival prediction," Bioinformatics, vol. 34, no. 14, pp. i446–i454, 2018.
[10] R. Argelaguet et al., "Multi-omics factor analysis—A framework for unsupervised integration of multi-omics data sets," Molecular Systems Biology, vol. 14, no. 6, 2018.

Biographies



Pallavi Dilip Waghmare received the Bachelor's Degree in Computer Engineering from Sanjivani College of Engineering Kopargaon in 2020 and the Master's Degree in Computer Engineering from Amrutvahaini College of Engineering in 2022. She is currently pursuing PhD from PCET's Pimpari Chinchwad University, Pune and working as an Assistant Professor at the Department of Computer Engineering, Faculty of Engineering, SVIT, Chincholi.



Mr. Vijay N. Patil received the Bachelor's Degree in Electronics Engineering from Shivaji University W.I.T. Solapur in 1992, the Master's Degree in Electronics Engineering from B.A. Marathwada University(Govt. College, Aurangabad) in 2000 and the PhD(Electronics Engineering) from North Maharashtra University, Jalgaon. He is currently working as Dean, School of Engineering and Technology, and PhD Program Director, PCET's Pimpari Chinchwad University, Pune

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

SenseBin: IoT-Based Solar Powered Smart Dustbin for Real-Time Waste Level, Gas Smell, and Lid Status Detection

Shree Ambekar[1], Prof. Sayali Belhe[1], Suyash Ambi[1], Akash Kumar[1], Sankalp Akkar[1]

Vishwakarma Institute of Technology, Pune, India

Email: shreeambekar09@gmail.com, sayali.belhe@vit.edu, suyashambi1@gmail.com, akashkumar1011k@gmail.com, sankalpakkar@gmail.com

Abstract.

Rapid urbanization has increased solid waste, making timely collection difficult. This paper presents SenseBin, a solar-powered IoT smart bin with separate dry and wet waste compartments. Ultrasonic, MQ135, and lid sensors monitor fill level, gas conditions, and lid status. Data are sent to Firebase and displayed on a web dashboard. Testing showed effective real-time monitoring, making SenseBin suitable for smart-city waste management.

Keywords. IoT; smart waste-bin; waste management; gas sensing; solar system; real-time monitoring.

1. INTRODUCTION

Increasing urban waste creates collection and hygiene challenges [1]. Traditional bins lack real-time monitoring of fill level, gas, and lid status, causing delays [2,3]. SenseBin addresses these issues using ultrasonic, MQ135, and lid sensors with solar-battery power and cloud-based monitoring.

2. RELATED WORK

Zanella et al. [1] introduced IoT-based smart-city services, while Arebey et al. [2] used RFID, GPS, GPRS, and GIS for bin tracking. Navghane et al. [3] proposed ultrasonic-based smart bins. LoRaWAN and cloud-based systems were explored by Silva et al. [4] and Chen et al. [5]. Other studies [6–9] highlighted IoT, gas sensing, wireless networks, and energy-autonomous waste monitoring. Solar-powered and wireless IoT systems [10–12] further support the feasibility of outdoor smart-bin solutions. Table 1 compares these systems based on fill level, gas sensing, lid status, and solar power.

Table 1. Comparison of SenseBin with representative prior smart-bin systems.

Work	Fill-level sensing	Gas sensing	Lid status	Solar powered
[2]	Yes (RFID/GPS)	No	No	No
[3]	Yes	No	No	No
[4]	Yes (LoRa)	No	No	No
[5]	Yes	Limited	No	No
This work	Yes (Ultrasonic)	Yes (MQ135)	Yes	Yes

3. PROPOSED SOLUTION

System Overview

SenseBin is a dual smart-dustbin system for separately monitoring dry and wet waste. Each bin houses ultrasonic, gas and lid sensors connected to its own ESP controller. As shown in

Figure 1, the system has four parts: a power-supply unit (solar panel, charge controller, battery), the smart dustbin unit (sensors and controller), a cloud server, and a user interface comprising a web dashboard and mobile app for remote monitoring of fill level, gas condition, lid position and battery status.

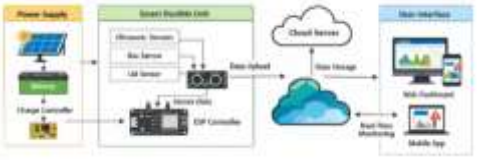


Figure 3.1. System architecture of the proposed SenseBin platform.

Working Principle

decomposition gases and the lid sensor monitors opening and closing. The ESP32 processes all readings and sends them to the cloud database for real-time dashboard monitoring.

4. HARDWARE ARCHITECTURE

Each bin uses an ESP32 [13] with ultrasonic, MQ135, lid, and battery sensors for real-time monitoring. Solar power provides energy, while battery voltage is monitored for power status [14].

5. SOFTWARE ARCHITECTURE

The firmware collects and transmits sensor data to Firebase [15]. The web dashboard displays fill level, gas, lid, and battery status in real time.



Figure 5.1. Data and control flow of the SenseBin smart waste-management system.

6. MATHEMATICAL MODELING AND CALCULATIONS

Fill Level Calculation

With H being the height of the internal bin (cm) and d, the distance measured by the ultrasonic sensor (cm), the is given by (1):

$$\textbf{\textit{Fill Level}}\ (\%) = \frac{H-d}{H} \times 100 \quad \ldots\ldots (1)$$

Once a threshold has been met (e.g. bin is 80% full), it can be collected.
Battery Percentage Calculation
Let Vm be the battery voltage at a current time and Vmin / Vmax, the safe minimum / maximum battery voltages. The battery capacity is calculated from (2):

$$\boldsymbol{Battery\ Percentage}\ (\%) = \frac{V_m - V_{min}}{V_{max} - V_{min}} \times \mathbf{100} \quad \text{..... (2)}$$

Gas Level Threshold Logic
If the measurements of the MQ135 sensor G are higher than a certain threshold Gthreshold, the air is considered unsafe, as expressed in (3):

$$\boldsymbol{If\ G > G_{threshold},\ Gas\ Condition = Unsafe} \quad \text{..... (3)}$$

The full classification used by SenseBin is summarized in Table 2.

Table 2. Gas-level threshold classification using the MQ135 sensor.

Sensor Value Range	Gas Level Classification	System Status
0 – 150	Safe	Normal
151 – 300	Moderate	Warning
Above 300	Toxic	Unsafe

7. RESULTS

The performance of SenseBin was tested under real-life conditions. The unit continuously logged the fill level, gas quality, lid status and battery charge and fed data to the dashboard in real time with no appreciable latency. Sensor readings were consistent for different fill levels and conditions. Observations on sensing unit are summarised in Table 3.

Table 3. Observations from SenseBin sensing units.

Parameter	Sensor Used	Observation
Waste fill level	Ultrasonic sensor	Accurate, consistent distance measurement
Lid status	Push button	Pressed = closed; not pressed = open
Gas presence	MQ135 gas sensor	Clear response to decomposition gases
Gas classification	MQ135 gas sensor	Reliable safe / unsafe identification
Dashboard update	ESP + cloud	Real-time updates without delay

Figure 7.1 shows the hardware implementation of SenseBin, including the ESP32, connected sensors, power-management circuitry, and the assembled dual dry/wet waste-bin prototype.



Figure 7.1 SenseBin Circuit, Physical Prototype, and Web Dashboard.

8. DISCUSSIONS

Low-cost sensors, ESP32, and Firebase enabled effective real-time monitoring of fill level, gas, and lid status, improving on conventional systems [2–5]. Solar-battery power supports energy-independent operation [10–12]. Limitations include the MQ135's non-specific gas detection and the prototype's small capacity.

9. CONCLUSIONS

SenseBin provides solar-powered, real-time monitoring of waste level, gas, and lid status using ultrasonic, MQ135, and lid sensors with ESP32 and Firebase. Future work includes selective gas detection, large-scale deployment, and route optimization.

10. REFERENCES

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[2] M. Arebey, M. A. Hannan, H. Basri, R. A. Begum, and H. Abdullah, "Integrated technologies for solid waste bin monitoring system," Environmental Monitoring and Assessment, vol. 177, no. 1–4, pp. 399–408, Jun. 2011.

[3] S. S. Navghane, M. S. Killedar, and V. M. Rohokale, "IoT based smart garbage and waste collection bin," International Journal of Advanced Research in Electronics and Communication Engineering, vol. 5, no. 5, pp. 1576–1578, May 2016.

[4] R. Silva, J. J. P. C. Rodrigues, and A. M. Alberti, "LoRaWAN-based smart waste management system," Sensors, vol. 18, no. 8, pp. 1–15, Aug. 2018.

[5] H. C. Chen, Y. C. Chen, and C. Y. Lin, "Development of an IoT-based smart waste management system," IEEE Access, vol. 8, pp. 123–132, 2020.

[6] M. Al-Maaded, N. Madi, A. Hodzic, and N. Ozerkan, "An overview of solid waste management and plastic recycling in Qatar," Journal of Polymers and the Environment, vol. 20, no. 1, pp. 186–194, Mar. 2012.

[7] S. Longhi et al., "Solid waste management architecture using wireless sensor network technology," in Proc. 5th Int. Conf. New Technologies, Mobility and Security (NTMS), May 2012, pp. 1–5.

[8] P. S. Pandey and V. Mahajan, "Smart waste management system using IoT: A review," International Journal of Engineering Research and Technology, vol. 8, no. 6, pp. 227–230, Jun. 2019.

[9] T. Anagnostopoulos et al., "Challenges and opportunities of waste management in IoT-enabled smart cities: A survey," IEEE Transactions on Sustainable Computing, vol. 2, no. 3, pp. 275–289, Jul.–Sep. 2017.

[10] M. K. Hasan, M. A. Hossain, and M. A. Rahman, "Solar-powered IoT-based smart monitoring system," International Journal of Renewable Energy Research, vol. 9, no. 4, pp. 1870–1878, Dec. 2019.

[11] C. Gomez and J. Paradells, "Wireless home automation networks: A survey of architectures and technologies," IEEE Communications Magazine, vol. 48, no. 6, pp. 92–101, Jun. 2010.

[12] J. Gutiérrez, J. F. Villa-Medina, A. Nieto-Garibay, and M. Á. Porta-Gándara, "Automated irrigation system using a wireless sensor network and GPRS module," IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 1, pp. 166–176, Jan. 2014.

[13] Espressif Systems, ESP32 Technical Reference Manual, Espressif Inc., Shanghai, China, 2023.

[14] Hanwei Electronics Co., MQ-135 Semiconductor Sensor for Air Quality Datasheet, Hanwei Electronics, Zhengzhou, China, 2022.
[15] Google, “Firebase Realtime Database,” Firebase Documentation, 2024. [Online]. Available: https://firebase.google.com/docs/database

Biographies



Shree Ambekar completed his HSC from D. G. Ruparel College, Mumbai, Maharashtra. He is currently pursuing a B.Tech. in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra, along with a B.Sc. in Data Science from IIT Madras.



Prof. Sayali Belhe is an Assistant Professor in AI and Data Science at Vishwakarma Institute of Technology, Pune, and is pursuing her Ph.D.



Suyash Ambi is pursuing a B.Tech. in Artificial Intelligence and Data Science at Vishwakarma Institute of Technology, Pune.



Akash Kumar is pursuing a B.Tech. in Artificial Intelligence and Data Science at Vishwakarma Institute of Technology, Pune.



Sankalp Akkar completed his HSC from M. S. Kothari Junior College, Nashik, Maharashtra. He is currently pursuing a B.Tech. in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Agentic AI Framework for Adaptive and Inclusive Tax Filing in India

Jay Nagose, Soniya Mate, Prajakta Kale, Dakshit Vardekar, Shantanu Bhavsar, Snehal Hon, Yogesh S. Deshmukh

Department of Information Technology, Sanjivani College of Engineering, Kopargaon, India, nagosejayraj2005@gmail.com, soniyamate10@gmail.com, prajaktakale2009@gmail.com, dakshit06@gmail.com, shantanubhavsar36@gmail.com, snehalhon10@gmail.com, yogeshdeshmukhit@sanjivani.org.in

Abstract.

The existing Indian tax preparation systems have not effectively addressed the technological diversity of the 90 million tax payers and are associated with 34% filing errors rate with low digital literacy and disabled population. The paper offers an Adaptive Agentic AI solution, that focusing on Tax Act 196 with reconsidered form-based filing as a dynamic human AI dialogue. An agent has four elements Tax Logic Agent, User Profiling Agent , Interaction Agent and validation. These work together through knowledge graph retrieval to streamline processes without compromising the statutory accuracy. The interface will change the complexity of navigation, typography, depth of explanation, and modalities in response to real-time user behavior and cognitive load. It is implemented with Vue.js frontend and FastAPI backend and QloRA fine, tuned LLM on more than 3200 tax sections, focusing on privacy through tokenization and secure computation. The assessment of 60 participants shows that, vulnerable user makes 67% lower errors.

Keywords. agentic systems, agent AI, oracle problem mitigation, tax compliance, Indian Income Tax Act, knowledge graph RAG, Llama fine-tuning, cognitive load adaptation, the inclusive design.

1. INTRODUCTION

1.1. The Indian Tax Compliance Crisis

Direct taxation of India is based, on the Income Tax Act, 1961, which has more than 90 million taxpayers and has become a tangled network of provisions, regulations and amendments [1]. There are provisions such as the Section 80 deductions and other provisions that require complex calculations and cross references, such as the capital gains Sections 45-55 [2] that overwhelm the average users. Research indicates the overall 34% of returns have problem to the statutory misinterpretation not data entry which causes, legal simplification is necessary.

Although 90% of the online filing had adopted, standard interfaces on the governments, and commercial websites will assume standard financial literacy and digital skills [3]. This

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non-Commercial 4.0 International

leaves out seniors, low-literacy users, and disabled persons, who will find it hard to deal with forms and opaque outputs that will easy. The oracle problem does undermine trust: users are not able to confirm results unless they have domain knowledge, which propagates mistakes and distrust [4]. The existing systems are form-based high cognitive load, non-adaptive [5], and non-transparent rule-based experts are not de-signed to customize to the needs of users [6]. None adapt dynamically complexity, explanations, and modes, marginalizing various taxpayers who need assistance [7].

1.2. The Adaptive Agentic Paradigm

This paper presents the Adaptive Agentic AI Framework in which the transforms or converts merely filing for interactive dialogues. Response times, errors, and queries, which are used as real-time cognitive profiling, allow making adaptations in terms of navigation, density, and depth. A multi agent system Tax Logic, User Profiling, Interaction, and Compliance Agents, works together, using knowledge graphs to reason relational law and provide explainable results [8]. This stack has a Vue.js UI with responsive and adaptive interfaces; FastAPI microservices and agent orchestration; a QLoRA-fine-tuned LLM [9] with hybrid retrieval to reduce hallucinations. Assessment through user experiments validates improved usability, 67 percent reduced errors, and increased trust, opening the way to a more inclusive tax compliance [10].

2. LITERATURE REVIEW

The literature shows that India's complex tax system and rigid platforms can increase filing errors. Although adaptive interfaces, knowledge graphs, LLMs, and multi-agent systems offer potential solutions, they are rarely integrated with inclusive tax software [11,12]. The proposed framework bridges this gap by combining adaptive interfaces, statutory grounding, multi-agent processing, and inclusive design.

3. KEY CONTRIBUTION

The proposed framework combines real-time adaptive interfaces, AI-enabled knowledge graphs, and multi-agent architecture to improve accessibility, statutory grounding, citation accuracy, and transparency in tax filing. The fine-tuned Llama 3.2 model achieved 51.5% accuracy, 51.8% precision, 51.5% recall, 51.6% F1-score, and 0.52 AUC, providing a baseline for future improvement using larger datasets and ensemble methods.

3.1. Knowledge Graph Retrieval

The KG-enhanced retrieval of the system glitters, Llama 3.2 with the baseline achieves 51.5% citation accuracy and 48.5% hallucination. Standard RAG increased to 71.3% accuracy 16.8% hallucination. Neo4j KG-RAG boosted it to 88.7%, 5.1%, and citations verified finalized at 94.2% 2.3%. Graph traversal is critical in statutory grounding in cross-referenced areas of law.

Table 2: Retrieval Pipeline Performance

Pipeline Stage	Citation Accuracy	Hallucination Rate	Response Time
Baseline Llama 3.2	51.5%	48.5%	0.8 s
Standard RAG	71.3%	16.8%	1.4 s

KG-RAG (Neo4j)	88.7%	5.1%	1.9 s
Citation Verifier	94.2%	2.3%	2.3 s

3.2. User Study Results

The adaptive interface reduced errors across all user groups, with vulnerable users showing the largest improvement, from 44.2% to 14.6% (67% reduction). This demonstrates the benefit of real-time adaptation for users facing literacy, complexity, or accessibility challenges.

Table 3: Error Rate Reduction and SUS (System Usability Scale) Score Comparison

User Group	Static Error Rate	Adaptive Error Rate	Static SUS*	Adaptive SUS*	Change	Improvement
Tech-Savvy	8.2%	4.1%	72.1	74.8	+2.7	50%
Average	21.3%	9.8%	58.3	68.2	+9.9	54%
Vulnerable	44.2%	14.6%	41.2	63.5	+22.3	67%

System Usability Scale (SUS) scores improved significantly, with vulnerable users increasing by 22.3 points to 63.5, while average users also showed improvement. Trust increased from 2.1 to 4.7 among vulnerable users and from 3.7 to 5.3 overall, indicating that clear explanations improved user confidence.

3.3. Implementation Validation

The prototype was implemented using Vue.js 3, FastAPI, Redis, Neo4j, FAISS, AutoGen, and Llama 3.2 QLoRA. It achieved 2.3 s P95 latency, supported 10,000 concurrent users, provided 99.7% citation coverage, and maintained 12 ms tokenization overhead.

3.4. Overall Result

Overall, the framework reduced filing errors and improved usability, trust, and inclusiveness. Its main strength lies in combining adaptive interaction, knowledge-graph retrieval, and transparent reasoning.

4. CONCLUSIONS

The proposed Adaptive Agentic AI Framework achieved 94.2% accuracy, reduced tax-filing errors by 67%, and improved usability by 22.3%. Despite limitations such as AUC = 0.52 and limited edge-case coverage, future work will focus on larger datasets, multilingual support, and GST and corporate compliance.

5. REFERENCES

[1] Sharma and A. Gupta," Challenges in Indian income tax compliance: A study of filing errors and digital literacy barriers," Journal of Financial Regulation and Compliance, vol. 29, no. 3, pp. 210–225, 2021.

[2] Mehta," Digital governance and tax administration in India: Opportunities and challenges under the Income Tax Act 1961," Indian Journal of Public Administration, vol. 67, no. 2, pp. 145–160, 2021.

[3] K. Dey and G. D. Abowd," Toward a better understanding of context and context awareness," Proceedings of the CHI Workshop on the What, Who, Where, When, and How of Context Awareness, 2000.
[4] Sweller," Cognitive load theory and educational technology," Educational Technology Research and Development, vol. 68, no. 1, pp. 1–16, 2020.
[5] Todi, B. Bailly, and L. Luyten," Familiarization: Restructuring layouts with explainable adaptive UI," ACM CHI Conference on Human Factors in Computing Systems, pp. 1–12, 2021.
[6] Wooldridge, An Introduction to Multiagent Systems, 2nd ed. Chichester, UK: Wiley, 2009.
[7] Park et al.," Generative agents: Interactive simulacra of human behavior," Proceedings of the 36th ACM UIST Symposium, pp. 1–22,2023.
[8] Lewis et al.," Retrieval-augmented generation for knowledge-intensive NLP tasks," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 9459–9474, 2020.
[9] Dettmers et al.," QLoRA: Efficient finetuning of quantized LLMs," Advances in Neural Information Processing Systems (NeurIPS), vol. 36, 2023.
[10] Touvron et al.," Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, 2023.
[11] Li et al.," A survey on legal judgment prediction with deep learning," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 4, pp. 1494–1507, 2022.
[12] Lewthwaite and D. Sloan," Exploring accessibility as a site of professional knowledge for web developers and UX designers," ACM W4A Web for All Conference, pp. 1–4, 2016.

Biographies



Jay Nagose received the bachelor's degree in Information Technology from Sanjivani College of Engineering in 2026. Research include LLMs, AI systems, RAG, ML & RL.



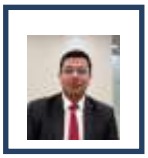
Soniya Mate received the bachelor's degree in Information Technology from Sanjivani College of Engineering in 2026. Her research interests include AI, ML, hyperspectral imaging Research on skin cancer detection using hyperspectral imaging deep learning.



Prajakta Kale received her B.Tech. in Information Technology from Sanjivani College of Engineering in 2026. Her interests include AI, ML, intelligent systems, and medical image analysis.



Dakshit Vardekar received the bachelor's degree in Information Technology from Sanjivani College of Engineering in 2026. His research areas include cloud computing and infrastructure design, system architecture.



Shantanu Bhavsar received the bachelor's degree in Information Technology from Sanjivani College of Engineering in 2026. his research areas include Human-Computer Interaction (HCI), and Accessibility Evaluation for Vulnerable Populations



Snehal Hon received the bachelor's degree in Information Technology from Sanjivani College of Engineering in 2026. Research interests include AI, data intelligence, ML.



Dr. Yogesh S. Deshmukh holds B.Tech., M.Tech., and Ph.D. degrees in Information Technology and Computer Science and Engineering. His research interests include AI and ML, with 20 Scopus-indexed publications.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Post-Quantum Cryptography for Securing Banking API: A Survey

Kshitij Dhake, Atharva Dhake, Kartik Tichkule, Abhinay Dhamankar & Sachin Pande

Department of Information Technology, SCTR's Pune Institute of Computer Technology, Pune, India, kdhake289@gmail.com, atharvadhake01@gmail.com, kartik.tichkule.11@gmail.com, agdhamankar@pict.edu, sspande@pict.edu

Abstract.

Recent strides in quantum computing have posed a challenge to conventional classical cryptographic algorithms like RSA and Elliptic Curve Cryptography (ECC). With the progression of quantum computing, there is a possibility of vulnerability to traditional cryptography in the future, leading to significant concerns from the perspective of the security of the banking industry. In this survey, we will review the adoption of Post-Quantum Cryptography (PQC), focusing on specific schemes like CRYSTALS-Kyber or CRYSTALS-Dilithium, together with an approach involving hybrid cryptosystems, ensuring seamless implementation. In addition, several challenges related to computational costs, difficulties of integration, and the growing significance of flexible deployment of cryptography will be considered. In summary, the objective of the research is to help financial institutions and scholars determine efficient quantum-safe security systems that offer protection for banking APIs.

1. INTRODUCTION

The modern banking system relies significantly on APIs and operates on a digital platform through different functions like financial transactions, payments, authentication, and interaction. At present, these interactions are protected using conventional public key cryptography techniques. Public key cryptography has been quite dependable so far because of the very complex nature of mathematical calculations that cannot be solved easily by conventional means.

Progress in quantum computing is raising significant questions about the future effectiveness of standard encryption techniques. Algorithms like Shor's have the potential to compromise traditional public-key cryptography much faster than current computers. While fully functional quantum computers have yet to be released, experts warn that sensitive financial information encrypted today could be at risk once these powerful machines become a reality.

"A delay like that makes banks more vulnerable to quantum computing algorithms down the road because banks handle confidential financial information."

The rise of quantum technology has created an urgent need to safeguard financial data. To address this, researchers and standardization bodies are working on post-quantum

cryptography (PQC) algorithms designed to protect against threats from both classical and quantum computers. This survey reviews how NIST-approved PQC algorithms are being integrated into the security systems of banking institutions and evaluates their compatibility with TLS, OAuth 2.0, JWT, and mTLS protocols.

2. RELATED WORKS

Extensive research has explored the effects of quantum computing on cryptographic systems employed in financial and banking sectors. Standard cryptographic mechanisms are widely used for securing online banking and financial communications. However, these systems are vulnerable to compromise through quantum attacks leveraging Shor's algorithm.

Bettale, Oliveira & Dottax (2022) discuss post-quantum cryptography systems specifically designed for banking systems and how quantum-resistant protocols can be integrated into banking infrastructure. Ramamoorthy & Singh (2025) examine how quantum algorithms affect classical public-key cryptography and how PQC can replace traditional schemes. Gujjala (2023) presents a new multi-factor authentication scheme for online banking that employs post-quantum cryptography.

Existing research primarily addresses architectural aspects or develops systems using PQC to improve authentication. However, little research has specifically investigated the implementation of NIST-standardized PQC algorithms in banking API protocols such as TLS, OAuth 2.0, JWT, and mTLS. This survey bridges that gap.

3. CONTRIBUTIONS

This survey makes the following contributions:

(1) Examines quantum security risks to banking APIs that use RSA and ECC cryptosystems, providing justification for adopting quantum-safe cryptography.

(2) Discusses primary categories of PQC, focusing on security and performance of NIST-standardized schemes: Key encapsulation using ML-KEM, digital signing through ML-DSA and hash based authentication with SLH-DSA.

(3) Assessing the applicability of Post-Quantum Cryptography (PQC) for securing bank APIs via TLS, OAuth 2.0, JSON Web Tokens (JWT), and mTLS.

(4) Hybrid model analysis and their potential for making an enabling easy transmission to quantum-resistant cryptography.

4. EVALUATION METHODOLOGY

This section integrates the current literature and available benchmark data. No new experimental results are presented here. The purpose is to discuss the quantum threat model, evaluate the key categories of PQC algorithms, and evaluate the viability of PQC for banking API security.

4.1. Quantum threat model

Shor's Algorithm: provides a quick way to factorize large integers and calculate discrete logarithms, thus rendering RSA and Elliptic Curve Cryptography encryption algorithms

insecure. The RSA-2048 algorithm is just one of many that may be considered a serious threat to security in future years.

Grover's Algorithm: is superior to brute force search, resulting in reduced security levels for symmetric algorithms. AES-256 must be used instead for quantum-resistant security.

Harvest Now, Decrypt Later: involves the interception of ciphertexts currently and decryption in the future when quantum computers are more advanced. This attack model shows the importance of implementing PQC quickly to protect critical data.

4.2. Standards for Cryptography

As per NIST, there are three standards for Post-Quantum Cryptography. This safeguards against both classical and quantum architectures for attacks:

ML-KEM (FIPS 203): A key encapsulation mechanism utilizing a lattice-based cryptography algorithm, originating from the CRYSTALS-Kyber KEM. ML-DSA (FIPS 204) and SLH-DSA (FIPS 205): A digital signature algorithm.

These PQC standards will be implemented through the OQS initiative in transport layer security and application programming interface.

The Kyber and Dilithium cryptosystem standards provide authentication to protect OAuth 2.0, JSON Web Tokens (JWT), and mTLs.

4.3. Analysis of PQC algorithms and comparison

Table 1 gives a comparative study of major post quantum cryptographic algorithms with respect to security foundation, key size, performance and suitability for banking applications.

Property	ML-KEM-768	ML-DSA-65	FALCON-1024	SLH-DSA (L3)
Purpose	KEM	Signature	Signature	Signature
Cryptography Type	Lattice (MLWE)	Lattice (MLWE)	NTRU Lattice	Hash-Based
Public Key size	1.2 KB	1.9 KB	1.8 KB	48 B
Ciphertext size	1.1 KB	3.3 KB	1.2 KB	16 KB
Operation time	60–90 µs	60–100 µs	100 µs	5–10 ms
Fit for Bank API	Excellent	Excellent	Moderate	Backup

TABLE 1. COMPARATIVE ANALYSIS OF PQC ALGORITHMS FOR BANKING APPLICATIONS.

Table 2 summarizes recommended PQC algorithm usage across various banking and financial applications based on performance and security characteristics.

Application	Main Algorithm Used	Backup / Alternative
TLS & Payment Gateways	ML-KEM-768 for secure key exchange	—
API Authentication	ML-DSA-65 for verifying identities	—
Digital Signatures	ML-DSA-65 as the primary signing method	SLH-DSA as a fallback option
Blockchain & High-Frequency Trading	FALCON-1024 for fast and secure signatures	ML-DSA-65 as backup
IoT & Embedded Devices	ML-DSA-44 optimized for lightweight devices	ML-KEM-512 as an alternative

TABLE 2. RECOMMENDED PQC ALGORITHM USAGE ACROSS BANKING APPLICATIONS.

4.4. PQC implementations in banking APIs

Hybrid methods combine Kyber and Dilithium with traditional cryptography for quantum resistance and backward compatibility. Hybrid protocols such as ECDHE + Kyber and ECDSA + Dilithium support secure API migration. PQC integration must address larger keys, computational overhead, and infrastructure compatibility. OQS and OpenSSL-OQS provide open-source tools for PQC and hybrid TLS integration.

5. DISCUSSIONS

Lattice-based schemes offer a good balance of security and efficiency for banking APIs. Classic McEliece provides strong security but has large keys, while SLH-DSA and FALCON suit specific requirements. Hybrid RSA/ECC + ML-KEM/ML-DSA approaches enable practical migration while supporting existing TLS, OAuth 2.0, JWT, and mTLS protocols.

6. CONCLUSION

Quantum computing threatens RSA and ECC, making PQC essential for banking API security. Lattice-based schemes are promising, while hybrid approaches support interoperability with existing systems. Future work should evaluate side-channel security, HSM performance, and low-latency financial applications.

7. REFERENCES

[1] Bettale, L., De Oliveira, M. & Dottax, E. 2022. Post-quantum protocols for banking applications. *In 4th PQC conference Proceding.*

[2] Gujjala, P.K.R. 2023. Quantum-enhanced multi-factor authentication framework for digital banking systems: a post-quantum cryptographic approach. *(IJMR)* 5(6).

[3] National Institute of Standards and Technology (NIST). 2024. *Module-Lattice-Based Key-Encapsulation Mechanism Standard (ML-KEM), FIPS 203; Module-Lattice-Based Digital Signature Standard (ML-DSA), FIPS 204; Stateless Hash-Based Digital Signature Standard (SLH-DSA), FIPS 205*. Gaithersburg, MD: NIST.

[4] Oleti, C.S. 2024. Post-quantum cryptographic architecture for secure banking: lattice-based implementation with blockchain integration. *(IJMR)* 6(2).

[5] Pasupuleti, M.K. 2025. Post-quantum cryptography: algorithms and implementation challenges. *International Journal of Academic and Industrial Research Innovations* 5(6).

[6] Ramamoorthy, U. & Singh, S. 2025. Post-quantum cryptography to secure future IT and financial systems. *IJARIIE* 11(4): 3788–3792.

[7] Rane, S.R., Shelake, V., Katte, T.B., Patil, S.V., Patil, D.V. & Musale, V. 2025. Quantum-resistant cryptographic algorithms: a comparative analysis for securing next-generation communication networks. *Journal of Information Systems Engineering and Management* 10(13s).

Biographies



Kshitij Dhake. Student, Department of Information Technology, SCTR's Pune Institute of Computer Technology, Pune, India



Atharva Dhake. Student, Department of Information Technology, SCTR's Pune Institute of Computer Technology, Pune, India



Kartik Tichkule. Student, Department of Information Technology, SCTR's Pune Institute of Computer Technology, Pune, India



Abhinay Dhamankar. Assistant Professor, Department of Information Technology, SCTR's Pune Institute of Computer Technology, Pune, India



Sachin Pande. Assistant Professor, Department of Information Technology, SCTR's Pune Institute of Computer Technology, Pune, India

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

AI-Driven Road Safety Monitoring and Civic Issue Reporting System Using Computer Vision

M. Venkata Soodarshan, Ketan G. Nibandhe, Vaibhav S. Nagawade, Dr. Chetan N. Aher

Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India, m.venkatsoodarshan@gmail.com*, nibandheketan2354*@gmail.com*, vaibhavnagawade2612*@gmail.com*, chetan.aher07*@gmail.com

Abstract.

The deterioration of road infrastructure affects traffic safety, vehicle maintenance, and transportation efficiency. Timely identification of potholes and estimation of their severity can support road maintenance and prioritization. Although deep learning methods can achieve high detection accuracy, they generally require large labeled datasets and substantial computational resources, which can limit their use in real-time and resource-constrained environments. This paper proposes a lightweight machine-learning pipeline for classifying pothole severity into Low, Medium, and High classes. Road images are resized and preprocessed, after which structural and surface characteristics are represented using Canny edge density and RGB color histograms. The resulting features are evaluated with Random Forest, Support Vector Machine (SVM), and XGBoost classifiers. On a dataset of 665 pothole images, Random Forest achieves the best performance, with 92.6% accuracy, 92.8% precision, 92.6% recall, and 92.5% F1-score. The results indicate that handcrafted image features combined with supervised learning can provide a practical alternative for lightweight road-condition monitoring and intelligent transportation applications.

Keywords. Road damage detection; pothole classification; machine learning; feature extraction; Random Forest.

1. INTRODUCTION

Urbanisation and increasing traffic cause road damage, creating safety risks, vehicle damage, and maintenance costs. Manual inspection is time-consuming and error-prone, motivating automated vision-based monitoring [1,2]. Although deep learning provides effective detection, its computational demands limit low-cost deployment [3–6]. This study uses Canny edge density and RGB histograms with Random Forest, SVM, and XGBoost to classify pothole severity for lightweight, real-time monitoring [7–9].

2. RELATED WORK

Early road-damage studies used image processing techniques such as enhancement, segmentation, edge detection, color, and texture features. Canny edge detection is widely used for effective edge localization [8], while image-processing and color-histogram methods support feature extraction [9,10]. However, illumination, shadows, and noise remain challenges [7].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Machine-learning methods such as SVM and Random Forest improve classification using handcrafted features [2]. CNN and transfer-learning approaches offer higher accuracy but require greater computational resources [3,4,6,11]. Therefore, this work proposes a compact feature-based pipeline for pothole severity classification.

3. DATASET AND PROPOSED METHOD

3.1. Dataset

The dataset contains 665 road images depicting potholes under varied road, lighting, and back-ground conditions. Potholes are represented by bounding-box annotations in Pascal VOC XML format. The dataset contains one damage category (pothole), while severity is assigned to three levels: Low, Medium, and High. Because the original data do not provide predefined severity labels, samples were manually categorized using pothole size, depth, and surface irregularity with consistent labeling criteria. Images are resized to 64 × 64 pixels and represented in RGB format. An 80:20 stratified train-test split is used so that the severity distribution is maintained across both sets.

3.2. System Overview

The proposed pipeline consists of image acquisition, preprocessing, feature extraction, super-vised classification, and severity prediction. The design intentionally uses lightweight operations so that inference can be performed on standard CPU-based systems.

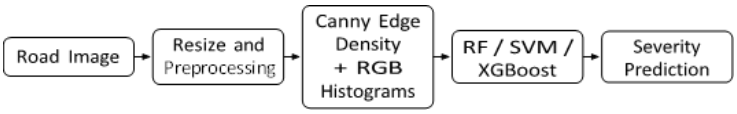


Figure 3.1.1: Lightweight road-damage severity classification pipeline.

3.3. Image Preprocessing

All images are resized to 64 × 64 pixels to reduce computational cost and provide a consis-tent input size. Grayscale images are used for edge-based processing, while the original RGB channels are retained for color analysis. Pixel values are normalized to reduce the influence of illumination differences. Canny edge detection is then applied to highlight structural discontinuities caused by potholes and other surface irregularities. The detector uses lower and upper thresholds of 50 and 150, respectively. Edge density is calculated as the ratio of edge pixels to total image pixels, producing a compact feature that represents the amount of structural variation in the road surface [8].

3.4. RGB Color Histogram Features

Structural information alone may not fully describe a damaged road surface. Therefore, RGB color information is incorporated through histogram analysis. Each RGB channel is quantized into eight bins, producing 8×8×8 = 512 color combinations. The resulting histogram represents the distribution of surface colors and can capture changes associated with asphalt wear, material degradation, water accumulation, and illumination [10]. The color features are combined with edge density to form the final feature vector.

3.5. Machine Learning Classification

The extracted features are evaluated using three supervised classifiers. Random Forest is an ensemble of decision trees and is used for its robustness and ability to model nonlinear relationships. A linear SVM is included as a margin-based classifier suitable for high-dimensional feature spaces. XGBoost is used as a gradient-boosted decision-tree method that can provide strong predictive performance with appropriate regularization. These models represent different learning strategies and provide a compact comparison of severity classification performance.

The classifiers are trained on the hand-labeled dataset. Random Forest uses 150 decision trees. Performance is measured using accuracy, precision, recall, and weighted F1-score. During inference, the same preprocessing and feature-extraction steps are applied to an unseen image before the trained model predicts its severity class.

4. EXPERIMENTAL RESULTS

4.1. Experimental Setup

The experiments were implemented in Python. OpenCV was used for image preprocessing and feature extraction, while scikit-learn and XGBoost were used for model training and evaluation. Experiments were performed on a standard CPU-based system without specialized GPU hardware. The three severity classes were Low, Medium, and High, and the dataset was divided using an 80:20 stratified split.

Table 1: Performance comparison of machine-learning models.

Model	Accuracy	Precision	Recall	F1-score
SVM (Linear)	88.4	88.7	88.4	88.2
XGBoost	90.9	91.1	90.9	90.7
Random Forest	92.6	92.8	92.6	92.5

The results in Table 1 show that Random Forest provides the best overall performance among the evaluated classifiers. It achieves 92.6% accuracy and maintains closely matched precision, recall, and F1-score values. XGBoost follows with 90.9% accuracy, while linear SVM achieves 88.4%. The results suggest that the nonlinear ensemble representation of Random Forest is well suited to the combined edge and color features used in this work.

5. DISCUSSION

The proposed Canny edge density and RGB histogram features provide effective, low-cost pothole severity classification. Random Forest performed best, while the method remains suitable for CPU and edge devices. Limitations include environmental variations and the small 665-image dataset.

6. CONCLUSION

The framework achieved 92.6% accuracy, 92.8% precision, 92.6% recall, and 92.5% F1-score using Random Forest. It offers a lightweight solution for real-time pothole severity monitoring.

7. REFERENCES

[1] N. Ma, J. Fan, W. Wang, J. Wu, Y. Jiang, L. Xie, and R. Fan, 'Computer vision for road imaging and pothole detection: A state-of-the-art review', *Transportation Safety and Environment*, vol. 4, no. 4, 2022, doi: 10.1093/tse/tdac026.

[2] V. Kaushik and B. S. Kalyan, 'Pothole detection system: A review of different methods used for detection', in *Proc. Int. Conf. on Computer Science, Engineering and Applications (ICCSEA)*, 2022, doi: 10.1109/ICCSEA54677.2022.9936360.

[3] M. Jakubec, E. Lieskovská, B. Bučko, and K. Zábovská, 'Comparison of CNN-based models for pothole detection in real-world adverse conditions', *Applied Sciences*, vol. 13, no. 9, p. 5810, 2023, doi: 10.3390/app13095810.

[4] R. Bharat, A. Ikotun, A. Ezugwu, L. Abualigah, M. Shehab, and R. Abu Zitar, 'A real-time automatic pothole detection system using convolution neural networks', *Applied and Computational Engineering*, vol. 6, no. 1, pp. 879–886, 2023.

[5] T. Tamagusko and A. Ferreira, 'Optimizing pothole detection in pavements: A comparative analysis of deep learning models', *Engineering Proceedings*, vol. 36, no. 1, p. 11, 2023, doi: 10.3390/engproc2023036011.

[6] K. A. Vinodhini and K. R. A. Sidhaarth, 'Pothole detection in bituminous road using CNN with transfer learning', *Measurement: Sensors*, vol. 31, p. 100940, 2024, doi: 10.1016/j.measen.2023.100940.

[7] H. B. Giri, H. Biradar, N. Gupta, P. V. Jirali, and J. N. Hemalatha, 'Pothole detection using image processing and data logging', in *Proc. 7th Int. Conf. on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, 2023, doi: 10.1109/CSITSS60515.2023.10334176.

[8] J. Canny, 'A computational approach to edge detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2008.

[10] M. J. Swain and D. H. Ballard, 'Color indexing', *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[11] M. Nalawi, M. Baghdadi, B. Alyateem, Z. S. Abo-Hammour, and A. Alsharkawi, 'Design of a real-time detection system for potholes and bumps using deep learning', in *Proc. Int. Conf. on Research and Education in Mechatronics (REM)*, 2024, pp. 160–165, doi: 10.1109/REM63063.2024.10735479.

Biographies



M. Venkat is a final-year B.Tech student in Computer Engineering at AISSMS Institute of Information Technology, Pune. His areas of interest include Artificial Intelligence, Machine Learning, Data Science, and Cloud Computing. He is interested in developing practical AI-based solutions and exploring emerging technologies.



Ketan G. Nibandhe is a Computer Engineering graduate from AISSMS Institute of Information Technology, Pune. His interests include machine learning, computer vision, and intelligent systems for real-world applications. His work focuses on developing lightweight and practical AI solutions for road safety and intelligent transportation



Vaibhav Nagawade is a final-year B.Tech student in Computer Engineering at AISSMS Institute of Information Technology, Pune. His areas of interest include Artificial Intelligence, Machine Learning, Data Science, and Software Development. He is interested in applying modern technologies to develop innovative and practical solutions



Dr. Chetan Nimba Aher is an Assistant Professor in the Computer Engineering Department at the All-India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, Maharashtra, India. With a Ph.D. in Computer Engineering from Kalinga Institute of Industrial Technology, Dr. Aher has 10 years of teaching experience. He has authored numerous research papers in SCI/SCOPUS indexed journals and actively participates in international conferences.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Leveraging Seaborn and XGBoost for Customer Analysis and Early Dissatisfaction Detection

Prof. Sayali Belhe, Siddhesh Kulthe, Chirayu Patil, Krishna Patil, Rushikesh Pathare, Dnyaneshwar Patil

Department of AI & DS, Vishwakarma Institute of Technology, Pune,
Sayali.belhe@vit.edu, siddheshkulte43@gmail.com, chirayupatil333@gmail.com,
krishna.patil24@vit.edu, rushikesh.pathare24@vit.edu, dnyaneshwar.patil24@vit.edu

Abstract.

Customer dissatisfaction can significantly affect business survival and purchase decisions. Traditional quantitative surveys measure satisfaction but often fail to capture customers' emotional responses and deeper attitudes. This study analyzes factors influencing purchase decisions using Seaborn for data analysis and XGBoost for prediction, aiming to identify customer dissatisfaction at an early stage and support timely corrective actions.

Keywords. Machine learning, Extreme Gradient Boosting, Customer Satisfaction.

1. INTRODUCTION

A customer is an individual or organization receiving goods or services. Identifying target customers supports effective marketing and customer-focused products (5, 8). XGBoost is an ensemble algorithm that combines decision trees to provide accurate predictions while handling missing data and reducing overfitting through regularization. The prediction is:

$\hat{y}_i = \varphi(x_i) = \Sigma_{k=1}^{K} f_k(x_i)$

Here, $\hat{y}_i$ is the predicted output, K is the number of trees, and $f_k(x_i)$ is the prediction of the k-th tree. XGBoost builds trees sequentially to correct previous errors and uses L1/L2 regularization and pruning to reduce overfitting. It also supports feature-importance analysis and parallel computing. Seaborn is a Python visualization library that works with Pandas to create clear statistical plots for distributions, correlations, and other relationships.

2. RELATED WORKS

2.1. Literature Review

Machine learning and visualization techniques, particularly Seaborn and XGBoost, are increasingly used for customer analysis and early dissatisfaction detection.

2.2. Seaborn In Customer Analysis

Seaborn helps visualize customer behaviour, preferences, distributions, correlations, and demographic patterns. Its integration with Pandas supports clear and effective data exploration [14].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2.3. XGBoost for Early Dissatisfaction Detection

XGBoost provides accurate prediction and identifies subtle behavioural patterns associated with customer dissatisfaction. Regularization and tree pruning improve generalization and reduce overfitting [15].

2.4. Integration Of Seaborn and XGBoost

Seaborn helps identify important customer features, while XGBoost uses these features for dissatisfaction prediction. Their integration provides both visual insights and predictive analysis of customer behaviour.

3. OBJECTIVES & SCOPE

3.1. Objective:

To identify factors such as price, features, promotions, and brand image that influence purchase decisions and early customer dissatisfaction.

3.2. Scope:

The study focuses on predicting early dissatisfaction and understanding how key customer and marketing factors influence purchase decisions for a new business without historical customer data.

4. PROCEDURE

The research examines customer preferences and perceptions using data collected through a simulated business environment. Participants interact with the products and respond, enabling analysis of purchasing behaviour.

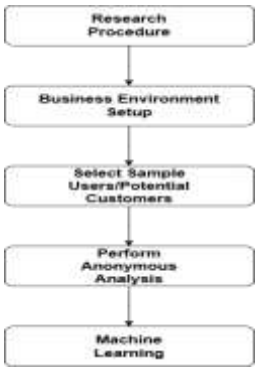


Figure 4.1: Research procedure

4.1. Business Environment Creation:

A simulated start-up business was created without existing customer relationships to study the effects of key variables on customer perceptions and purchasing decisions [6].

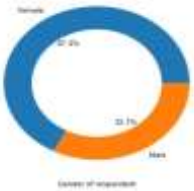


Figure 4.1.1: Gender of Respondents

The sample consisted of 67.3% female and 32.7% male respondents, indicating higher female participation.

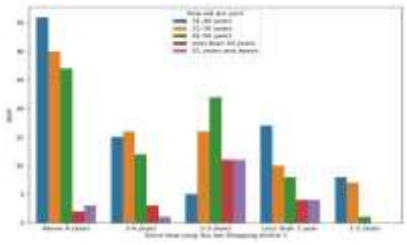


Figure 4.1.2: Shopping and Online Purchase Experience by Age
The 21–30 age group represented the largest proportion of respondents with shopping and online purchasing experience, reflecting their familiarity with digital technologies.



Figure 4.1.3: City of Purchase

The figure presents the geographical distribution of respondents who purchased the product.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental results provide insights into customer purchasing behaviour and the factors influencing purchase decisions. The analysis examines purchase frequency, cart abandonment, shopping duration, and important product features. These relationships help identify patterns of customer engagement and disengagement within the simulated business environment.

6. REFERENCES

[1] R. L. Oliver, 'A cognitive model of the antecedents and consequences of satisfaction decisions,' J. Marketing Res., vol. 17, no. 4, pp. 460–469, Nov. 1980.

[2] E. Garbarino and M. S. Johnson, ‘The different roles of satisfaction, trust, and commitment in customer relationships,’ J. Marketing, vol. 63, no. 2, pp. 70–87, Apr. 1999.
[3] D. Gupta, M. Gupta, R. K. Dhanaraj, B. Balusamy, and P. M. Gupta, Eds., Computational Intelligence for Autonomous Finance. Hoboken, NJ: John Wiley & Sons, 2024.
[4] D. Kang and Y. Park, ‘Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach,’ Expert Syst. Appl., vol. 41, no. 4, pp. 1041–1050.
[5] Bob E. Hayes, ‘Measuring Customer Satisfaction and Loyalty,’ 2008.
[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, ‘The Elements of Statistical Learning,’ 2009.
[7] Stephanie D. H. Evergreen, ‘Effective Data Visualization: The Right Chart for the Right Data,’ 2016.
[8] Bob E. Hayes, ‘Customer Satisfaction: The Customer Experience Through the Customer’s Eyes,’ 1998.
[9] D. P. Verma, D. D. Gupta, and D. M. Gupta, “A comprehensive study on the implications of ESG performance disclosure in the promotion of sustainable development by firms,” 2024.
[10] C. Ranaweera and J. Prabhu, ‘The influence of satisfaction, trust and switching barriers on customer retention in a continuous purchasing setting,’ Int. J. …

Biographies



Prof. Sayali Belhe received the bachelor's degree in Computer Science and Engineering (CSE) and the master's degree in Information Technology (IT). She is currently pursuing a PhD and working as a faculty member in the field of Computer Science and Engineering.



Siddhesh Kulthe is an Artificial Intelligence and Data Science (AI&DS) Engineer currently working in the IT industry. He is an industry expert with expertise in the field of Information Technology.



Chirayu Patil currently pursuing a bachelor's degree in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune. He is currently in his third year of study**.**



Krishna Patil is currently pursuing a bachelor's degree in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune. He is currently in his third year of study.



Rushikesh Pathare is currently pursuing a bachelor's degree in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune. He is currently in his third year of study.



Dnyaneshwar Patil is currently pursuing a bachelor's degree in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune. He is currently in his third year of study.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

CropTalk- AI-Driven Crop Disease Prediction And Management System

Aparna Torade[1], Apurv Deshmukh[2], Harshad Jadhav[3], Monali Tingane[4]

[1],[2],[3],[4],School of Computing, MIT Art Design and Technology University, Pune, Maharashtra, 412201, India, aparna.torade@mituniversity.edu.in

Abstract.

There is a lack of easily accessible knowledge-sharing platforms in agriculture due to recent technology breakthroughs. We have created a communal online application for the agricultural community that is similar to Twitter in order to address this. Through direct messaging (DMs), the website enables farmers to solve each other's issues, share success stories, and spread expertise. The software also incorporates a real-time weather API, which offers crucial meteorological information for agricultural planning, for even more user control. We have also added a machine learning model that can forecast any plant disease using our knowledge of data science and artificial intelligence.

Keywords. Machine Learning, Plant-Disease Prediction, Direct Messaging (DMs)

1. INTRODUCTION

Agriculture plays a vital role in the economy, but farmers often lack timely access to information and peer support. This work proposes a digital farming platform that enables farmer communication through direct messaging, provides real-time weather information through an API, and uses a machine-learning model to identify plant diseases from leaf images. The platform aims to create an integrated ecosystem for agricultural information, disease detection, and farmer collaboration.

2. RELATED WORK

CNN-based models have demonstrated strong performance in plant disease detection, with reported accuracies above 99% using datasets such as PlantVillage [1,4,7,8]. However, these approaches often require large datasets and significant computational resources. Traditional machine-learning methods, including Random Forest, SVM, and ensemble models, have also shown potential for early disease prediction [6].

Digital platforms and social media can improve knowledge sharing, collaboration, and agricultural decision-making among farmers [3,9,11]. However, challenges such as poor connectivity, limited infrastructure, and internet costs can restrict their use [9]. Existing studies largely address disease detection or farmer communication separately; therefore, this work combines plant disease detection, weather information, and farmer networking within a single platform.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. KEY CONTRIBUTION

Through the development of a single, farmer focused system that combines prediction of crop diseases and sharing knowledge via community, the Croptalk system is able to improve agriculture through innovations in technology. Unlike other methods which address these three different issues separately, the Croptalk system combines them all into one convenient application using machine learning, social networking features, and current weather data. Realistic implementation is stressed throughout, as well as scalability and simplicity for farmers who lack technical skills. Such a system will enable more effective farming activities and decisions.

4. METHOD

To ensure reliability, scalability, and usability, the proposed system for disease detection and agricultural information dissemination was designed using a structured methodology involving different phases. System designing, system development, deployment, testing, and system support were some of the phases involved in the process. First, there was the design of the system that involved both functional and non-functional requirements of the system. Afterward, there was the selection of a cloud infrastructure that ensured high scalability and availability. This resulted in the decision to choose a cloud platform that had back-end hosting, database management, and API integration. In this regard, the chosen cloud provider is Amazon Web Services. There was the setting up of the front-end and back-end development environments that involved installing such key software as Postman (for API testing) and Git (for version control).

There was the designing of the frontend using JavaScript, HTML/CSS, and React. This helped in ensuring a sophisticated interface. This is done to ensure a sophisticated interface through which farmers can submit images of crops affected by diseases. The CNN-based machine learning algorithm was developed and included as one of the modules of the system. Both normal and disease-affected crop images, which were collected from an open-source database, were used to train this module. Pre-processing techniques were also applied to the images to improve the performance of the model under different conditions. After training, the algorithm was then used as a disease detection API service in real-time. On the other hand, the frontend application was successfully deployed in connection with the backend APIs, while the backend was hosted on the AWS cloud platform during deployment. To ensure secure connections to the database and APIs as well as proper authentication process, environment variables were established.

Then came the testing process of the system. In this case, integration testing was done to ensure that there were no problems in the communication between various system modules, while functional testing was performed to verify the performance of every system component. In addition, performance testing was conducted to ensure that the system could function efficiently and securely. Security testing was also done. Detected bugs were fixed before the system release.

Pilot testing was done by implementing the application with a few numbers of users to get inputs from actual use. For guiding users in using different features such as uploading images, getting disease diagnosis, and accessing information about agriculture, training and

manuals were provided. Modifications were done for increasing the effectiveness and usability of the application, depending on pilot testing feedbacks.

After implementation, the system is continuously monitored to ensure that there will be no problem with its working and technical assistance can be readily provided. Feedback methods were embedded in the system to help with continuous learning, improve machine learning, and ultimately extend the scope of the system to other types of diseases and crops.

5. EXPERIMENTAL RESULTS

The functional aspects, accuracy, and performance of the Crop talk system were analysed based on practical experiments conducted on the system. The outcome is a clear picture of how the system integrates three thing's that farmers are looking for, in one seamless platform: Crop disease detection, community communication and weather awareness. To test the social side of things a group of users was given access: farmers talking to other farmers to send messages, post updates and interact in real-time. It was found to be very durable; communication was relatively swift and users were able to share information and react to one another without any difficulties. Not only was it a test for the feature, it was more like a community rallying together. Testing the weather module weather information that counts when it counts where real data was collected from a variety of sources and tested for successful reception. Always accurate and relevant information on temperature, humidity and forecasts. The API could provide data within reasonable time for farmers to not have to wait for the data when making a time-sensitive decision. A module that detects diseases, evaluated using the various leaf images healthy and diseased. CNN model has shown to be effective in diagnostic of diseases with accuracy and timeliness. Most significantly, it was of practical use: farmers could have an idea of what was ailing their plants and use it without going to an expert first. Finally, the functional and integration testing were performed to validate the system's reliability when used as a whole.

6. DISCUSSIONS

The platform really helps to make communication between farmers easier; farmers can now update one another, ask questions, and share their knowledge, without having to call or pass the message by mouth. It's the quick and casual conversation you'd have face-to-face but before! Weather Insights Instead of trying to guess today or getting a road weather forecast, farmers have access to real-time local data and it allows them to plan around it. Which translates into less surprises and better judgement on when to plant, irrigate or harvest. Before the disease can spread, the CNN model won't get it right in all conditions and image quality is important too but it will give farmers a first impression quickly and practically, of what may be wrong with their crops.

7. CONCLUSIONS

Farmer will be needing tools that carry disease prediction and identification, sharing information and statistics collected in one place, not distributed across detached place. The system was made to fill that gap. The aim was to take honestly valuable technologies and fit them around the way farmers actually work rather than trying to fit agricultural needs into a technical term.To do that the system combines a CNN-based model to notice and analyze

the crop diseases, live weather data from external APIs and a social layer for farmers to share experiences and guidance. All services run as a web application with React for the interface and Node.js with Mongo DB for the cloud backend. Farmers on the platform might be able to detect disease faster, make better decisions with weather context, and connect with others facing the same challenges. The system has its limitations the model's precision can vary depending on image quality and environmental conditions, and it's worth being upfront about that. Future scope may include investigate on evolving wider variety of datasets and supporting several crops and languages.

8. REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016, doi: 10.3389/fpls.2016.01419. https://doi.org/10.3389/fpls.2016.01419.

[2] K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009. https://doi.org/10.1016/j.compag.2018.01.009

[3] J. Sharp, E. Imerman, and G. Peters, "Community Supported Agriculture (CSA): Building Social Connections," Journal of Community Development Society, vol. 33, no. 1, pp. 76–91, 2002, doi: 10.1080/15575330209490145. https://doi.org/10.1080/15575330209490145

[4] K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009. https://doi.org/10.1016/j.compag.2018.01.009

[5] O. Kulkarni, "Crop Disease Detection Using Deep Learning," in Proc. Int. Conf. Advances in Computing, Communication, and Control (ICAC3), 2021, doi: 10.1109/ICAC353642.2021.9697166.
https://doi.org/10.1109/ICAC353642.2021.9697 166

[6] E. P. Pathrose and P. Verma, "E-commerce and teleworking: Business opportunities in the COVID-19 era," Electrochemical Society Transactions, vol. 107, no. 1, pp. 12585–12594, 2022.

[7] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019, doi: 10.1016/j.compag.2018.03.032. https://doi.org/10.1016/j.compag.2018.03.032

[8] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016, doi: 10.3389/fpls.2016.01419. https://doi.org/10.3389/fpls.2016.01419

[9] C. W. Kuria, "Use of Social Media as a Source of Agricultural Information by Smallholder Farmers: A Case Study of Lower Kabete, Kiambu County," Master's thesis, University of Nairobi, Nairobi, Kenya, 2014.

[10] S. Burbi and K. Hartless Rose, "The Role of Internet and Social Media in the Diffusion of Knowledge and Innovation Among Farmers," Centre for Agroecology, Water and Resilience, Coventry University, 2016.

[11] P. Verma and R. Verma, "Predictive analytics for sustainable investment portfolios in banking," Multidisciplinary Science Journal, vol. 7, 2025, Art. no. 2025ss0231.

Biographies



Prof. Aparna Torade having 11 years of combined experience in the IT industry and academia. Currently working as an Assistant Professor at MIT Art, Design and Technology University, Loni Kalbhor, Pune. I am also pursuing my PhD at MIT ADT University, focusing on Blockchain and Federated Learning-based security. My areas of interest include cybersecurity, blockchain technology, federated learning, and data science.



Dr. Monali Tingane is an Assistant Professor in Computer Science and Engineering with over nine years of academic and industry experience. She holds a Ph.D. in Machine Learning from Sant Gadge Baba Amravati University.



Apurv Deshmukh Completed a Bachelor of Technology (B.Tech.) in Computer Science and Engineering (CSE – Core) from the School of Computing, MIT Art, Design and Technology University, Pune, Maharashtra, India



Harshad Jadhav Computer Science and Engineering (CSE – Core) graduate from the School of Computing, MIT Art, Design and Technology University,

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Collaborative Generative AI Agents for Enhanced Organizational Innovation

Prof. Komal Munde[1], Prof Suvarna Pawar[2], Prof. Kanchan Wankhade[3] , Prof. Sneha Deshmukh[4], Prof. Sachin Thanekar[5], Prof. Priya Khune[6]

[1,2,4,5,6]MIT Art, Design and Technology University, Loni Kalbhor Pune,3 VIT,Bibewadi Pune,

[1]komal.munde@mituniversity.edu.in,[2]suvarna.pawar@mituniversity.edu.in, [3]kanchan.wankhade@vit.edu,4sneha.deshmukh@mituniveristy.edu.in,5sachin.thanekar@ mituniversity.edu.in,6priya.khune@mituniversity.edu.in

Abstract.

This study proposes a Multi-Agent Generative AI Framework for organizational innovation. Multiple autonomous agents perform ideation, evaluation, domain reasoning, and strategy alignment through coordination and feedback loops. The framework reduces redundancy and bias while improving creativity, solution quality, scalability, and decision efficiency compared with single-agent and manual approaches.

Keywords. Multi-Agent Systems, Generative AI, Organizational Innovation, Autonomous Agents, AI Collaboration.

1. INTRODUCTION

Innovation is essential for organizational competitiveness, but traditional brainstorming and expert-based methods can suffer from bias, limited scalability, and knowledge silos [1,2]. Generative AI and LLMs support idea generation, reasoning, and decision-making, but single-agent systems often provide limited perspectives and repetitive solutions [3–6]. Multi-Agent Systems enable autonomous agents to collaborate and solve complex problems with greater diversity and scalability [7–9]. However, their application to organizational innovation remains limited [10]. This study proposes a Multi-Agent Generative AI Framework for collaborative ideation, evaluation, and strategic alignment.

2. RELATED WORK

2.1 Generative AI in Organizations

Generative AI enhances creativity, decision-making, automation, and productivity across organizations [11]. However, its adoption introduces challenges related to governance, trust, and knowledge sharing [13]. Existing applications mainly focus on task automation rather than collaborative innovation [9].

2.2 Multi Agent Systems (MAS)

Multi-Agent Systems use autonomous agents that cooperate or compete to solve complex problems through coordination and decentralized decision-making [4–7]. Integrating LLMs

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

further enhances agent reasoning and collaboration. However, existing research largely emphasizes technical coordination rather than organizational innovation.

***2.3 Bridging the Gap Between Theory and Sustainable Practice*:**

Table 1: Organizational Innovation Challenges and Framework Contribution[9]

Challenge	**Limitation**	**Multi-Agent Solution**
Cognitive bias	Limited diverse thinking	Diverse agents generate ideas
Knowledge silos	Fragmented information	Agents share cross-functional knowledge
Limited scalability	Difficult parallel tasks	Parallel agent execution
Weak strategic alignment	Goals may be missed	Goal-driven agents align outputs
Low transparency	Unclear AI reasoning	Explainable agent decisions
Slow decisions	Time-consuming consensus	Faster parallel decision-making

2.4 Research Gap

Although GenAI and MAS research has advanced, existing studies mainly focus on single-agent systems, task automation, or agent coordination [11–15]. Limited attention has been given to Multi-Agent GenAI frameworks for organizational innovation, particularly collaborative ideation, evaluation, decision-making, and strategic alignment.

3. Proposed Multi Agent Generative AI Framework

This section proposes Multi Agent Generative AI Framework of Organizational innovation which is a hybrid of generative artificial intelligence and Multi agent system (MAS) to support the organizational innovation practices.

3.1 Framework Overview

The proposed framework is a distributed, collaborative generative system that transforms organizational inputs into innovative, strategically aligned outputs. It follows an Input–Agent Collaboration–Output pipeline, where inter-agent interaction drives innovation rather than isolated generation [9].



Figure 3.1.1: Planned Multi Agent Generative AI Framework for Organizational Innovation

Organizational inputs are assigned to domain-specific LLM agents that collaborate through a coordination layer for task allocation and iterative refinement [2,3,14]. The framework generates and ranks ideas, solutions, and strategic recommendations, while its modular architecture supports scalability and flexible agent configuration [2].

4. Application to Organizational Innovation

The sector gives evidence to the usability of the Multi Agent generative AI framework proposed on representative organizational innovation use case, as well as a comparative study with the conventional and single agent approaches.

4.1 Use Case Scenario: New Product Development

Giving to the diagram, the Multi Agent Generative Framework of New Product Development (NPD) works in a efficient and repeated process that is aimed at converting the high-level statistics of an organization into sophisticated outputs of innovations. [6]



Figure 4.1.2 Application of the Proposed Framework to New Product Development

The framework follows four stages: Input Integration, Multi-Agent Generative Core, Coordination and Feedback, and Iterative Refinement, transforming organizational inputs into refined, innovative, and strategically aligned solutions.

4.2 Multi-Agent Workflow Stages and Corresponding Evaluation Matrices

Table 2: Innovation Lifecycle Stages and Evaluation Criteria [4]

Challenge	Limitation	Multi-Agent Solution
Bias in thinking	Repetitive ideas	Diverse agents + feedback
Knowledge gaps	Isolated information	Collaborative agents
Scalability	Limited handling	Parallel execution
Goal misalignment	Not goal-driven	Goal-oriented agents
Low transparency	Unclear decisions	Explainable steps
Slow process	Time-consuming	Faster parallel decisions

4.3 Target Evaluation Metrics for Validating AI-Driven New Product Development

The providing Metrix begins a benchmarking framework to evaluate the efficiency of the **Multi Agent GenAI** model against traditional human and single agent approaches. [6]



Figure 4.3.3: Target Evaluation Metrics for Innovation Quality and Process Performance

6. Conclusion and Future Work

This paper presents a Multi-Agent Generative AI Framework that combines GenAI with multi-agent principles to improve organizational innovation through specialized creativity, analysis, evaluation, and planning agents [4,13]. Future work will focus on empirical evaluation, human-in-the-loop systems, governance, transparency, and ethical accountability. The framework can also be extended to cross-organizational innovation.

7. References

[1] M. Wooldridge, An Introduction to MultiAgent Systems, 2nd ed. Chichester, U.K.: Wiley, 2009.

[2] R. G. Cooper, Winning at New Products, 4th ed. New York, NY, USA: Basic Books, 2017.

[3] K. B. Kahn, "Understanding innovation," Business Horizons, vol. 61, no. 3, pp. 453–460, 2018.

[4] J. R. Wang et al., "Cooperative and competitive multi-agent systems: From optimization to games," IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 5, pp. 763–783, 2022.

[5] E. Mollick and L. Mollick, "Using AI to implement effective teaching and innovation strategies," Harvard Business School Working Paper, 2023.

[6] J. Park et al., "Generative agents: Interactive simulacra of human behavior," in Proc. International Conference on Machine Learning (ICML), 2023.

[7] E. Mollick, "What generative AI means for innovation," MIT Sloan Management Review, vol. 65, no. 1, pp. 1–6, 2023.

[8] J. Holmström, "How organizations can innovate with generative AI," ScienceDirect, 2024.

[9] X. Li et al., "A survey on LLM-based multi-agent systems," Springer, 2024.

[10] X. Li et al., "A survey on large language model based autonomous agents," Artificial Intelligence Review, 2024.

[11] K. V. Wankhade et al., "A meta-learner-integrated stacking voting ensemble network for cervical malignancy classification," in Proc. International Conference on Emerging Smart Computing and Informatics (ESCI), 2024.

[12] W. Jin et al., "A comprehensive survey on multi-agent cooperative decision-making: Scenarios, approaches, challenges and perspectives," arXiv, 2025.

[13] P. Pimpalkar and A. M. Jagtap, "Plant disease detection: PyramidNet-ICNN architecture with modified BIRCH segmentation," Journal of Phytopathology, vol. 173, no. 3, 2025.

Biographies



Prof.Komal Sunil Munde.She is currently working as an Assistant Professor at the Department of Computer Engineering, MITADT University. She has 9+ year experience. His research areas include Machine learning, deep learning.



Dr. Suvarna Pawae With 19+ years of teaching experience, I have served as a Head of Department for 15+ years, mentoring students and guiding UG/PG projects and research.I hold a Ph.D.



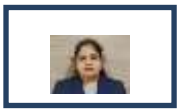
Prof. Kanchan V. Wankhade,is an academician and researcher. She working as a Assistant Professor at Vishwakarma Institute of Technology at Bibwewadi Pune. She is perusing Phd from Symbiosis Institute of Technology, Lavale Pune Her research interests include IoT, Data Science, Artificial Intelligence, Machine Learning and Generative AI.



Sneha Deshmukh is an Assistant Professor at MIT ADT University. with over 12 years of teaching experience. She is currently pursuing her Ph.D. and has research interests in Machine Learning, Artificial Intelligence, and Internet of Things (IoT).



Dr. Sachin Arun Thanekar is an academician and researcher. His research interests include Artificial Intelligence, Machine Learning, Natural Language Processing, Explainable AI, and Air Quality Prediction.



Prof. Khune P.A She is working as an Assistant Professor in MIT Art,Design and Technology University,Pune. She has 10+ years of teaching experience.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Smart and Automated Attendance Monitoring System

Monali Tingane[1], Rohan Patil[2]

[1,2]School of Computing, MIT Art, Design and Technology University, Pune, Maharashtra, 412201, India, monali.tingane@mituniversity.edu.in, patilrony48@gmail.com

Abstract.

One of the most basic but time-consuming activities in schools is attendance management. The suggested system also includes the practical additions of time locked attendance windows, location validation via a lightweight GPS, one-photo-per- lecture enforcement, face recognition validation by confidence, background similarity check, and randomized selfie prompts, eliminating the need to use live video streams to decrease proxy attendance. The attendance data and images of the classroom are safely stored in the cloud, and this allows transparency and verification by the faculty after the fact. The experimental results conclude that the system is suitable to be deployed on a large scale in academic institutions, as it was shown to be more accurate, has lower cases of attendance fraud, and is also more time saving than the traditional ones.

Keywords. Automated Attendance, Smart Attendance System, Facial Recognition, Artificial Intelligence

1. INTRODUCTION

Attendance administration is a crucial administrative activity in educational institutions because it bears direct implications for student involvement, discipline, and schoolwork involvement. This is despite its importance because attendance is usually taken by manual roll call, which not only wastes lecture time but also is subject to error and manipulation by human beings. To eliminate these shortcomings, schools have embraced semi automatic methods like RFID cards, Biometric scanners and mobile applications of attendance. Nonetheless, these systems bring about additional problems such as high cost of infrastructure, hygiene issues, device dependency and susceptibility to proxy attendance by sharing cards or spoofing. Even face recognition systems have a way of being abused unless proper validation mechanisms are put in place. The students can also seek to provide the images that have been precaptured, to attend remotely or take advantage of system loopholes. Thus, the optimal smart attendance system should be a combination of automation and realistic limitations that will promote authenticity without losing functionality. The present paper suggests the Smart Attendance System based on the Classroom Image Recognition that would be implemented on a mobile application. When a faculty takes a single classroom photograph during a predetermined lecture break, the system will record attendance without any video detection, thus removing the possibility of live video detection. In order to enhance authenticity, the system incorporates time-locked attendance windows, lightweight GPS-based location verification, one- photo per-lecture

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

enforcement, face recognition through confidence, similarity analysis of backgrounds and randomised selfie prompts. All images and attendance records of the classroom are safely placed on the cloud, which allows transparency and post-checking by faculty. The proposed Smart Attendance System provides an efficient, transparent, and robust solution to the traditional attendance systems.

2. RELATED WORK

A number of research works have been conducted on automated attendance systems based on alternative technologies to minimise manual labour and enhance precision. The RFID-based systems are quicker in payment of attendance, but they need specific hardware and are still susceptible to proxy attendance via card sharing [1]. Biometric systems (fingerprint or iris scan) are more effective in identity verification, but require a high cost of installation, hygiene issues, and cannot be expanded to large classrooms [2]. Face recognition-based attendance systems are becoming popular with the development of computer vision, since they are non-invasive. Conventional methods based on Haar Cascade and Local Binary Pattern Histogram (LBPH) are satisfactory in a controlled setting but fail in adverse lighting conditions, occlusions and pose changes [3]. Very recent deep learning approaches, such as CNNs, FaceNet and ArcFace, are more accurate and robust, but most of them use fixed camera systems or video streams, which raise their computation cost and raise privacy concerns [4], [5]. Attendance systems based on mobile phones with GPS or Wi-Fi are flexible and have fewer infrastructure demands, but can easily be spoofed by their location and abused by misusing a device [6]. In general, solutions that are in place have issues of proxy attendance, privacy and scalability. These shortcomings encourage the necessity to have a single-image, mobile-first attendance system with inherent anti proxy capabilities, which the proposed solution will assume.

3. PROBLEM DEFINITION AND MOTIVATION

3.1 Problem Statement

Attendance processes that are manual, e.g. roll calls, use up precious teaching hours and disrupt the normal teach-back process. Such techniques are also subject to the human factor, such as wrong marking and deliberate distortion. The larger the size of the classes, the less viable and easier to track attendance manually. Such challenges were met by hardware-based attendance systems such as RFID cards and biometric scanners that bring about new restrictions. RFID-based systems also involve extra infrastructure and are vulnerable to proxy attendance by sharing of cards. Biometric systems are more secure, but they are expensive to install and maintain hygienically, and they have problems of scalability, especially in large classrooms where the students are dense. The above restrictions are the reasons why a smart, reliable, and scalable attendance system is required, which reduces dependency on infrastructure but ensures accuracy, authenticity, and transparency.

3.2 Proposed System Overview

The Smart Attendance System proposed is a mobile first-based solution that can be used to automatically mark attendance by taking pictures of the classroom. The system works by enabling faculty members to take one picture of the classroom through a mobile application when a specific lecture time frame is within a preestablished period. The image acts as the key to the attendance processing, and as such, it does not require constant video monitoring or fixed cameras. After capturing the classroom image, an AI-based face detection and face recognition algorithm is applied to the picture to recognise the students attending the classroom. The recognised faces are matched with a student database that is previously registered, and attendance is automatically recorded depending on the confidence of the recognition. To improve credibility and eliminate misuse, the system will incorporate various verification functions, time limits, location checks and environment check consistency. All the registration data and other related images of classrooms are carefully saved in a cloud database. The faculty members will be given a verification dashboard where they can access attendance outcomes, deal with suspicious situations, and produce attendance reports. Faculty Dashboard of Verification and Reporting: Enables teachers to check on pending cases, fix wrong classification and produce attendance reports that can be used by the academic staff and administration.

3.3 Key Improvements and anty proxy mechanisms

The proposed system has a number of phone-based limits to curb the fraudulent practices and overcome the limitations of the current attendance systems. A. Time-Locked Attendance Window: Submission of attendance is only allowed within a preset lecture period, such as the first 10 minutes of the session. After this time window expires, the system will automatically block the marking of attendance. This system averts late submissions and remote proxy attempts to make the attendance a representation of the classroom attendance. B. Lightweight Location validation: The system confirms the position of the mobile device used by the faculty via GPS at the classroom capture image. The place must not exceed a given radius outside the classroom, or one is not permitted to attend. In order to support the principle of privacy, this strategy will maintain the physical presence despite the permanent location tracking, therefore reducing location-based fraud. C. One-Photo-Per- Lecture Rule: The system is one photo-per-lecture to deter the abuse of the system by submitting the same picture several times or swapping the image. The image of a classroom can be uploaded once per lecture session in order to have consistency and integrity in the attendance data. D. Face Recognition based on confidence: The face recognition module computes a score of similarity of each detected face, and this score is used to identify the attendance status: 90 per cent.: Automatically set to Present. 700-900: Considered as a waiting faculty review. < 70%: Marked as Absent. This hierarchical level of classification balances the position between 100 per cent automation and human analysis to enhance the overall reliability and reduce false positives. E. Background Analysis: Analysis of similarity. To make the system more realistic, background similarity checks are performed, and the visual features that can be compared include wall color, classroom boards and seating arrangements.

4. METHODOLOGY AND EXPERIMENTAL RESULTS

The faculty opens a session and takes a picture in a classroom during every lecture with a determination of a time period, which guarantees the availability of real time and valid attendance input without round the clock monitoring. The picture takes through the process of face detection on AI and then face recognition using deep learning, which detects all visible faces regardless of any changes in lighting or position and then compares them with the existing embeddings to produce similarity measures. With these scores attendance can be marked- high-confidence matches are marked present, borderline cases will be marked to review, and low-confidence matches will be marked absent, which further validation, such as location and background checks, will prevent proxy attendance. Lastly, all the information such as attendance records, images, and timestamps and logs are stored in a cloud-based platform allowing real-time accessibility, scalability, and automated report generation to be reviewed and integrated by the faculty with institutional systems. The system has been experimented with in the classroom setup with 20-50 students. The system has been experimented with in the classroom setup. The results indicate that under normal conditions of light, recognition is about 95 per cent. Within a few seconds, marking attendance was done. A major decrease in proxy attendance because of the combined vetting systems. Little faculty intervention was needed.

5. CONCLUSION AND FUTURE SCOPE

This paper has discussed a Smart Attendance System that uses classroom image recognition as a mobile app to make the process of marking attendance automatic. The system has a balance of accuracy, privacy and scalability by incorporating realistic anti-proxy technologies and evading live video processing. The positive effect of the research is confirmed by the improvement of efficiency and the decrease in attendance fraud. Future directions involve the improvement of recognition accuracy in the presence of occlusion, the incorporation of an attendance analytics dashboard, and a multi-class scheduler, as well as privacy-conserving AI methods.

6. REFERENCES

[1] S. Jain and R. Singh, “RFID Based Smart Attendance System,” International Journal of Computer Applications, vol. 178, no. 36, pp. 25–30, 2019.

[2] A. K. Sharma and P. Verma, “Face Recognition Based Attendance System Using Deep Learning,” in Proc. IEEE Int. Conf. on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 412–417.

[3] A. Patil, S. Kulkarni, and R. Deshmukh, “Automated Attendance System Using Face Recognition,” International Journal of Computer Science and Engineering, vol. 7, no. 5, pp. 45–50,2018.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 815–823.

[5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 4690–4699.

[6] R. Gupta, S. Bhattacharya, and P. Das, “Mobile- Based Attendance System Using GPS and Cloud Computing,” International Journal of Engineering and Technology, vol. 11, no. 2, pp. 75–82, 2019.
[7] K. Alhanaee, “Face Recognition Smart Attendance System Using Deep Learning,” Procedia Computer Science, vol. 187, pp. 123–130, 2021.
[8] R. Sawhney and R. Kacker, “Real-Time Smart Attendance System Using Face Recogni- tion,” Semantic Scholar, 2021.
[9] A. S. Lateef, “Facial Recognition Technology-Based Attendance System,” International Journal of Computer Science and Mobile Computing, vol. 12, no. 1, pp. 1–10, 2023.
[10] A. Jadhav, “Smart Attendance Monitoring System Using Biometric,” SSRN Electronic Journal, 2023.
[11] S. A. Shah and A. Abuzneid, “IoT-Based Smart Attendance System Using RFID and Cloud Computing,” International Journal of Computer Applications, vol. 176, no. 7, pp. 10–16, 2023.
[12] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human- Level Performance in Face Verification,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 2014, pp. 1701–1708.

Biographies



Dr. Monali Tingane is an Assistant Professor in Computer Science and Engineering with over nine years of academic and industry experience. She holds a Ph.D. in Machine Learning from Sant Gadge Baba Amravati University. Her research focuses on ML.



Viren Lahane is a Rohan Patil, a Computer Science Engineering graduate specializing in Artificial Intelligence and Analytics. My areas of interest include Artificial Intelligence, Machine Learning, Data Analytics, and developing innovative technology solutions.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Wellness Buddy: An AI-Enabled Emotion-Aware Companion for Digital Wellbeing

Supriya P. Jagtap, Arya Pawar, Samarth Bidkar, Nidhish Chincholkar, Aarush Sengupta and Adithya Sagri

Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India, supriya.jagtap@vit.edu, arya.pawar24@vit.edu, samarth.bidkar24@vit.edu, nidhish.chincholkar24@vit.edu, aarush.sengupta24@vit.edu, adithya.sagri24@vit.edu

Abstract.

Existing music recommenders mainly use listening history and collaborative filtering, overlooking users' current emotions. Wellness Buddy integrates real-time facial emotion recognition, emotion-aware music recommendation, and digital wellbeing tools. A CNN trained on FER2013 classifies four emotions, while a content-based recommender uses valence, energy, tempo, and acousticness for music selection. The system also includes Pomodoro sessions, hydration reminders, journaling, and an LLaMA-3.1 wellness assistant. It achieves 93% accuracy, 0.92 precision/recall/F1, and 40 ms (~25 FPS) latency, supporting real-time interaction.

Keywords. Emotion recognition, digital wellbeing, music recommendation, convolutional neural networks, human-computer interaction.

1. INTRODUCTION

Advances in AI and deep learning have enabled practical emotion-aware computing for personalized digital wellness and content delivery. However, conventional music recommenders mainly rely on listening history and overlook users' current emotions. Wellness Buddy addresses this gap through real-time facial emotion recognition and emotion-based music recommendation, complemented by Pomodoro timers, reminders, and an AI conversational assistant to support productivity and wellbeing.

2. RELATED WORK

Studies show that multimodal emotion recognition using facial, speech, and physiological data can improve accuracy [1,2,6,9]. CNN and deep-learning models achieve strong performance in emotion and music recommendation but often require high computational resources and large datasets [8,17]. Existing systems also face challenges such as limited modalities, cold-start issues, privacy, and real-world validation [4,9–16]. Therefore, a scalable and privacy-aware framework integrating multiple data sources is needed.

Table 1 compares existing approaches with the proposed system.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Ref	Authors & Year	Key Findings	Limitations / Research Gap
[1]	Kim et al. (2010)	Multimodal input improves prediction of music preference	Limited personalisation; no real-time adaptation
[2]	Soleymani et al. (2017)	Multimodal fusion improves emotion understanding	Needs scalable handling of large heterogeneous data
[3]	Pichappan (2025)	Emotion improves recommendation relevance	Cold-start and sparse user data
[4]	Yousefian Jazi et al. (2019)	Interaction data improves recommendations	Few features; no multimodal data
[6]	Chakravathi et al.	Facial + speech raises recognition accuracy	No large-scale validation or physiological data
[8]	Van den Oord et al.	Predicts latent preferences from audio content	Ignores user emotion and multimodal signals
[17]	Bhagat et al.	82.56% training / 65.68% validation accuracy	Low validation accuracy; emotion confusion
[9]	Abdullah et al.	Multimodal outperforms unimodal systems	Real-time processing; dataset standardisation
-	Proposed (Wellness Buddy)	Unified emotion detection, recommendation and wellbeing tools; 93% accuracy at 40 ms	Single modality at present; multimodal extension planned

3. KEY CONTRIBUTION

The proposed framework integrates real-time emotion recognition, emotion-aware music recommendation, productivity tools, and wellbeing support into a single system. It combines CNN-based emotion detection, valence- and energy-based recommendation, and an LLaMA-3.1 chatbot to provide personalized wellness and emotional support.

4. METHOD

The system uses Haar cascade and CNN-based emotion detection to classify four emotions from 48×48 facial images. Detected emotions guide music recommendation using valence, energy, acousticness, and tempo. Additional features include Pomodoro, hydration reminders, journaling, scheduling, and an LLaMA-3.1 chatbot.

5. EXPERIMENTS

Each module—emotion detection, music recommendation, productivity tools, and chatbot—was tested individually and as an integrated system for accuracy, functionality, and responsiveness.

6. RESULTS

The emotion classifier achieved 93% accuracy, with 0.92 precision, recall, and F1-score. The average latency was 40 ms (~25 FPS), enabling real-time interaction. The music

recommender provided emotion-matched tracks, while productivity and chatbot modules functioned correctly.

Table 2. Performance evaluation metrics of the proposed system.

Metric	Formula	Experimental Value
Accuracy	(TP + TN) / (TP + TN + FP + FN)	93%
Precision	TP / (TP + FP)	0.92
Recall	TP / (TP + FN)	0.92
F1-score	2PR / (P + R)	0.92
Total latency	T_face + T_class + T_update	40 ms
Theoretical FPS	1000 / T_total	25 FPS

As shown in Table 2, the system sustains high classification performance while keeping latency low enough for real-time use.

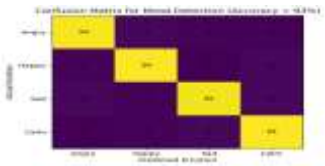


Fig.6.1. Confusion matrix for the four-class facial emotion recognition model.

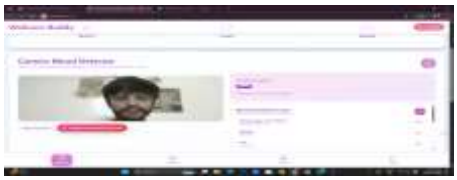


Fig. 6.2. Real-time emotion detection and music recommendation dashboard.

7. DISCUSSION

The proposed system provides accurate, real-time emotion recognition and emotion-based recommendations, supporting personalized wellness and productivity. However, performance may decline under poor lighting, occlusion, and subtle expressions due to its reliance on facial input.

8. CONCLUSIONS

Wellness Buddy combines CNN-based emotion recognition, emotion-aware music recommendation, productivity tools, and an LLaMA-3.1 chatbot. It achieved 93% accuracy, approximately 0.92 precision, recall, and F1-score, and 40 ms latency (~25 FPS). Future work will focus on multimodal inputs, advanced models, and personalized recommendations.

9. REFERENCES

[1] Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., Turnbull, D.: Music emotion recognition: a state-of-the-art review. In: Proc. 11th Int. Society for Music Information Retrieval Conf. (ISMIR), pp. 255-266 (2010).

[2] Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S.-F., Pantic, M.: A survey of multimodal sentiment analysis. Image and Vision Computing 65, 3-14 (2017).

[3] Pichappan, P.: A review of the emotion-induced music recommendation systems. Journal of Digital Information Management 23(2) (2025). ISSN 0972-7272.

[4] Yousefian Jazi, S., Kaedi, M., Fatemi, A.: An emotion-aware music recommender system: bridging the user's interaction and music recommendation. Multimedia Tools and Applications 80, 13559-13574 (2021).
[5] Singh, A.K., Mandal, D., Kumar, S., Venkat, D.: Emotion-based music recommendation system using deep learning. In: 2025 3rd IEEE Int. Conf. on Industrial Electronics: Developments & Applications (ICIDeA), Bhubaneswar, pp. 1-6 (2025).
[6] Chakravathi, K., Tejashwini, G., SriVibhavana, G., Shirisha, C.: Context-aware music recommendation using multimodal emotion recognition. Int. J. of Bioinformatics and Biomedical Engineering Research (IJBAR). ISSN 2249-3352.
[7] Talaghat, M.A., Parvinnia, E., Mehrabi, M., Boostani, R.: A multimodal deep network for music emotion recognition using audio chorus and lyrics. IEEE Access 13, 177509-177519 (2025).
[8] Van den Oord, A., Dieleman, S., Schrauwen, B.: Deep content-based music recommendation. In: Advances in Neural Information Processing Systems (NIPS) Workshop on Deep Learning and Representation Learning (2013).
[9] Saleem, S.M., Ameen, S.Y., Sadeeq, M.A.M., Zeebaree, S.R.M.: Multimodal emotion recognition using deep learning. Journal of Applied Science and Technology Trends 2(2), 52-58 (2021).
[10] Sunitha, M., Adilakshmi, T., Chitturi, S.S., Kapila, P.: Deep learning-based emotion-aware music recommendation system. In: Proc. 2nd Int. Conf. on Cognitive and Intelligent Computing (ICCIC 2022), Cognitive Science and Technology. Springer, Singapore (2023).
[11] Sathya, A.R., Varma, A.R.: Emotion-based song recommendation system. In: Intelligent Data Engineering and Analytics (FICTA 2023), Smart Innovation, Systems and Technologies, vol. 371. Springer, Singapore (2023).
[12] Regulwar, G.B., Kathirisetty, N.: Predictive music based on mood. Int. J. of Scientific Research in Science, Engineering and Technology (IJSRSET). ISSN 2395-1990.
[13] Mollahosseini, A., Hasani, B., Mahoor, M.H.: AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing 10(1), 18-31 (2019).
[14] Li, S., Deng, W.: Deep facial expression recognition: a survey. IEEE Transactions on Affective Computing 13(3), 1195-1215 (2022).
[15] Shah, D., Chavan, K., Shah, S., Kanani, P.: Real-time facial emotion recognition. In: 2021 2nd Global Conf. for Advancement in Technology (GCAT), Bangalore, pp. 1-4 (2021).
[16] Zhou, N., Liang, R., Shi, W.: A lightweight convolutional neural network for real-time facial expression detection. IEEE Access 9, 5573-5584 (2021).
[17] Bhagat, D., Vakil, A., Gupta, R.K., Kumar, A.: Facial emotion recognition (FER) using convolutional neural network (CNN). In: ICMLDE 2023, Procedia Computer Science 235, 2079-2089 (2024).

Biographies



Supriya P. Jagtap received the BE degree from SPPU in 2012 , ME degree from SPPU 2016 and Phd from Symbiosis International university in 2024



Arya Pawar Pursuing B.Tech degree from VIT Pune (TY)



Samarth Bidkar Pursuing B.Tech degree from VIT Pune (TY)



Nidhish Chincholkar Pursuing B.Tech degree from VIT Pune (TY)



Aarush Sengupta Pursuing B.Tech degree from VIT Pune (TY)



Adithya Sagri Pursuing B.Tech degree from VIT Pune (TY)

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Virtual Chemistry Lab

Priyanka Kamble[1], Soham Sane[2], Ayush Bhosale[3], Himanshu Macchi[4], Saish Mekal[5]

[1]Faculty, Department of Electronics Engineering, St. John College of Engineering and Management, Palghar, India, priyankaka@sjcem.edu.in

[2345]Student, Department of Computer Engineering, St. John College of Engineering and Management, Palghar, India, 122soham2122@sjcem.edu.in, 122ayush2123@sjcem.edu.in, 222himanshu2005@sjcem.edu.in, 122saish2118@sjcem.edu.in

Abstract

Virtual Reality (VR) has been increasingly adopted in chemistry education to address challenges such as visualization of abstract concepts, laboratory safety, accessibility, and student engagement. This paper presents the design and implementation of a VR-based chemistry laboratory and reviews related research on secondary and post-secondary learning. The system supports molecular structure visualization, organic reactions, titration experiments, and instrumentation techniques such as infrared spectrometry. Findings indicate that VR significantly improves spatial and visual understanding of sub-microscopic chemical phenomena, enhances procedural familiarity, and increases student motivation. However, VR is not always superior to traditional labs for conceptual mastery; learning outcomes depend strongly on instructional design, timing of VR introduction, and degree of physical interaction. Enhanced learning is observed when VR is integrated into laboratory sessions rather than used solely as a pre-lab tool. Overall, VR is a valuable complementary technology that improves visualization, accessibility, and practical learning when effectively combined with traditional chemistry instruction.

Keywords

Virtual Reality (VR), Chemistry Education, Virtual Chemistry Laboratory, Laboratory Simulation, Molecular Visualization, Abstract Chemistry Concepts, Organic Chemistry, Immersive Learning, Experimental Chemistry, Educational Technology.

1. INTRODUCTION

Chemistry is challenging because students must understand macroscopic, symbolic, and sub-microscopic concepts, often causing cognitive overload and misconceptions. Virtual Reality (VR) provides immersive 3D environments for visualizing molecular processes and practising laboratory procedures safely. Studies show that effective VR outcomes depend on instructional design, timing, and learner interaction. This paper reviews VR applications in chemistry, including titration, organic reactions, and infrared spectroscopy [1–3].

2. LITERATURE REVIEW

Recent reviews highlight the growing use of VR in chemistry education [4–6]. Virtual laboratories address the cost, safety, and accessibility limitations of traditional labs while helping students understand invisible molecular processes [7,8]. Studies also report

improved academic achievement among undergraduate chemistry students using virtual laboratories [14,19].

2.1. Visualization of Abstract Concepts

A principal strength of VR is the visualization of sub-microscopic phenomena. Students can interact directly with three-dimensional models of atoms and molecules, thereby improving spatial representation of molecular geometry and stoichiometry. Such immersive imaging is especially valuable in organic chemistry and titration simulations, reducing reliance on pure imagination [9]. Environments such as VICH-Lab further encourage independent learning.

2.2. Embodiment, Interaction and Procedural Training

Immersive VR links practical actions (manipulating virtual tools) with abstract concepts. Curriculum-based experiments indicate that while VR can yield superior representational learning, traditional physical equipment may still produce stronger conceptual outcomes in some cases [11]. Nevertheless, VR significantly raises self-efficacy and interest [10]. Procedural training—practising titration sequences or organic-reaction setups in a controlled virtual setting—reduces anxiety and builds confidence before students enter the wet laboratory [1, 2].

3. PROPOSED SYSTEM

The proposed Virtual Chemistry Lab (VCL) is conceived as a strategic complement, not a complete substitute, for conventional laboratories. It expands conceptual knowledge, procedural expertise and accessibility.

3.1. System Objectives

The principal objectives are: (i) to enhance visualization of abstract and sub-microscopic entities through immersive molecular models; (ii) to provide a safe environment for practising laboratory tasks and complex instrumentation; and (iii) to support integrated learning by embedding VR interventions within laboratory sessions rather than treating them as isolated pre-lab activities.

3.2. System Architecture

The architecture comprises five functional modules: (1) User Interface Module – navigation and height calibration; (2) Virtual Laboratory Environment – realistic laboratory layout; (3) Molecular Visualization Module – real-time depiction of sub-microscopic reactions; (4) Instructional Guidance and Feedback Module – automated guidance and adaptive feedback; and (5) Experiment Selection and Logging. Two primary scenes are implemented: a Main Menu Scene for height adjustment and module selection, and a Laboratory Scene that replicates a typical high-school chemistry laboratory complete with workbench, dynamic periodic table, instructional whiteboards, virtual sinks and emergency exits.

3.3. Virtual Environment and Molecular Visualization

The laboratory is designed to resemble a real-world setting, featuring benches, glassware, safety equipment and reagents. Core simulations cover organic reactions, titration and instrumentation-based analysis. Macroscopic changes (colour shifts, gas evolution) are

tightly coupled with molecular visualizations so that students can switch, at critical experimental phases, from the laboratory view to an atom-and-bond representation. This dual-scale approach reduces cognitive ambiguity while preserving immersion.

4. IMPLEMENTATION AND EXPERIMENTAL SETUP

VCL was developed on the Unity Real-Time Development Platform, chosen for its robust physics engine and extensive XR support. Interaction and chemical-behaviour logic were coded in C# within Microsoft Visual Studio. The system targets Oculus (Meta) Quest series headsets, which provide 6-Degrees-of-Freedom tracking and hand controllers that map realistic grasping, twisting and button interactions. Development and testing were performed on a VR-ready PC equipped with a dedicated GPU to maintain a stable frame rate of 72–90 FPS, thereby minimising motion sickness.

5. RESULTS AND DISCUSSION

The system provides a 3D VR chemistry laboratory with ray-casting controls, experimental equipment, a periodic table, and safety features. It supports four combination and decomposition reactions: $Al + I_2 \rightarrow AlI_3$, $CaO + H_2O \rightarrow Ca(OH)_2$, $CaCO_3 \rightarrow CaO + CO_2$, and $2FeSO_4 \rightarrow Fe_2O_3 + SO_3 + SO_2$. User feedback indicates improved spatial understanding, procedural confidence, and motivation, with greater benefits when VR is combined with hands-on laboratory activities [11].

6. CONCLUSION

VR provides an effective and engaging approach for teaching abstract and sub-microscopic chemistry concepts. The proposed system improves visualization and laboratory learning, although its effectiveness depends on appropriate instructional design and integration with traditional hands-on experiments.

7. FUTURE SCOPE

Future research should focus on hybrid learning platforms that combine real-world experiments with concurrent virtual visualizations, adaptive systems that personalise learning pathways through real-time performance data, and multi-user collaborative environments. Such developments are expected to maximise conceptual understanding, practical skill acquisition and overall learning efficiency.

8. ACKNOWLEDGEMENT

The authors thank the researchers and educators whose prior work formed the foundation of this study. Their contributions to pedagogical design, VR implementation strategies and learning outcomes in virtual laboratory environments have been invaluable. The authors also acknowledge the teachers and students who continue to explore progressive modes of chemistry instruction and thereby help create more accessible, stimulating and effective learning environments.

9. REFERENCES

[1] O. C. Agbonifo, O. A. Sarumi and Y. M. Akinola, "A Chemistry Laboratory Platform with Virtual Reality to support Adaptive Learning among students," Research in Learning Technology, vol. 28, pp. 1–9, Sep. 2020.

[2] K. Kounlaxay, D. Yao, M. W. Ha and S. K. Kim, "Design of Virtual Reality System in Organic Chemistry," Intelligent Automation and Soft Computing, vol. 30, no. 2, pp. 1–16, 2022.

[3] C. L. Dunnagan et al., "Production and Evaluation of a Realistic Immersive Virtual Reality Organic Chemistry Laboratory," J. Chem. Educ.

[4] R. M. Lopes et al., "Virtual Reality to Teach Students in Laboratories: A Bibliometric and Network Analysis," J. Chem. Educ., vol. 101, no. 2, pp. 501–513, Jan. 2024.

[5] P. Chan et al., "Virtual Chemistry Laboratory: A Systematic Literature Review," Computers and Education Open, vol. 2, 100024, 2021.

[6] T. N. M. Guruloo and K. Osman, Int. J. Educ., vol. 15, no. 4, p. 758, 2024.

[7] M. Morozov et al., "Virtual chemistry laboratory for school education," IEEE Int. Conf. on Advanced Learning Technologies, pp. 916–917, 2004.

[8] Z. Tatli and A. Ayas, "Virtual laboratory applications in chemistry education," Procedia – Social and Behavioral Sciences, vol. 9, pp. 1741–1745, 2010.

[9] N. Fitriyana et al., "Promoting Students' Learning Independence and Achievement by means of Virtual Chemistry Lab (VICH-Lab)," 2024.

[10] A. Gungor et al., "The Use of Virtual Reality in a Chemistry Lab and Its Effect on Self-Efficacy and Interest of Students," 2022.

[11] S. Qorbani, S. Dalili, A. Arya and C. Joslin, "Learning in immersive VR: Curriculum-based experiment in chemistry education," Educ. Sci., vol. 14, no. 5, p. 476, May 2024.

[12] J. L. Davenport et al., "Whether and How Authentic Contexts With a Virtual Chemistry Lab Support Learning," J. Chem. Educ., vol. 95, no. 7, pp. 1064–1070, 2018.

[13] H. Bazie, "The Effect of Virtual Laboratories on Academic Achievement of Undergraduate Students of Chemistry," 2024.

[14] E. Hu-Au and L. Lee, "Learning Abstract Chemistry Concepts using Virtual Reality," Electronics, vol. 13, no. 16, p. 3197, Aug. 2024.

Biographies



Saish Mekal is currently pursuing a bachelor's degree in Computer Engineering. His interests include full-stack web development, artificial intelligence, machine learning, and

educational technologies. He contributed to the development of **Virtual Chemistry Lab**, an interactive platform for digital chemistry experiments and simulations.



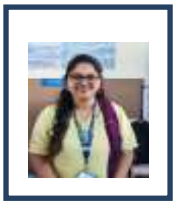
Ayush Bhosale received his Bachelor of Engineering degree in Computer Engineering from St. John College of Engineering and Management, Palghar. He worked on application development, architecture design, and system integration for the **Virtual Chemistry Lab** project.



Soham Sane received his Bachelor's degree in Computer Engineering from St. John College of Engineering and Management, Palghar. He handled literature survey, documentation, and presentation preparation for the **Virtual Chemistry Lab** project.



Himanshu Machhi received his Bachelor's degree in Computer Engineering from St. John College of Engineering and Management, Palghar. He handled literature survey, documentation, and presentation preparation for the **Virtual Chemistry Lab** project.



Priyanka Kamble is a faculty member in the Department of Computer Engineering at St. John College of Engineering and Management, Palghar. She served as the project guide for the **Virtual Chemistry** Lab project, providing academic guidance, technical insights, and support throughout the development and implementation of the project.

An Integrated Weather Intelligence Platform for Historical Analysis and Real-Time Extreme Event Alerts

Omkar Raskar [1], Prof. Sayali Belhe[1], Prathamesh Kumbhare[1],

Hitesh Pawar [1], Sameer Shaikh [1] , Sandesh Rabade [1]

[1] Vishwakarma Institute of Technology, Pune, India, sayali.belhe@vit.edu, omkar.raskar24@vit.edu, prathamesh.kumbhare24@vit.edu, hitesh.pawar24@vit.edu, sameer.shaikh24@vit.edu, sandesh.rabade24@vit.edu.

Abstract

Extreme weather events require integrated systems for data management, forecasting, and real-time alerts. This paper proposes an Integrated Weather Intelligence Platform combining multi-source meteorological data, machine-learning forecasting, and threshold-based extreme-event detection through a secure Flask/FastAPI backend. Using 52,560 hourly observations from multiple Indian weather stations, ARIMA, Random Forest, and LSTM achieved 82.5%, 87.3%, and 91.6% accuracy, respectively, with LSTM performing best. The alert system detected heat waves, heavy rainfall, storms, and cyclones with less than 3 seconds of notification latency, supporting timely decisions in agriculture, disaster management, transportation, and environmental research.

Keywords. Weather Data Portal; API; Meteorological Data; CSV Dataset; Backend Architecture.

1. INTRODUCTION

Weather intelligence supports agriculture, transportation, disaster management, and environmental research by enabling better planning and risk reduction. However, existing systems often suffer from fragmented data, poor accessibility, and limited analytics. This work proposes a Weather Data Portal with centralized data management, secure APIs, real-time access, and predictive analytics to improve weather monitoring and decision-making.

2. RELATED WORK

ARIMA performs well for linear forecasting but is limited with nonlinear weather patterns [11]. LSTM better captures long-term dependencies [4], while Weather2K supports multivariate forecasting [13]. The proposed platform combines LSTM, React, secure APIs, and real-time alerts to provide integrated weather intelligence.



Fig.2.1. Smart Weather Monitoring Architecture

3. METHOD

The integrated weather intelligence platform development design is designed in a multi-layered system that is used to process data high concurrency and make accurate predictions. There are three levels in this system, based on which they are: Data Layer to access data, Application Layer to apply business logic and ML analytics, and Presentation Layer to interact with the users

Table 1. System Overview Of The Proposed Weather & Climate Data Portal

Component	**Technology**	**Purpose**	**Component**
Data Acquisition	REST APIs	Data collection	Data Acquisition
Storage	CSV	Data storage	Storage
Processing	Python	Data preprocessing	Processing
Prediction	ML Models	Forecasting	Prediction
Alerts	Threshold Logic	Event detection	Alerts
Visualization	React, Chart.js	Data display	Visualization
Security	JWT, HTTPS	Data protection	Security

Experiment:

To demonstrate the proposed strategy, a series of experiments were conducted, and the integrity of data and the model preparedness were prioritized.

4. DATA COLLECTION AND PREPROCESSING:

It was decided to take the historical data which could be analyzed as a trend analyzer with the help of the Weather2K benchmark [13]. RESTful calls were made to weather APIs to receive information on the weather conditions in the atmosphere. The raw data was taken through a controlled preprocessing pipeline which included cleaning to eliminate some noise and anomalies. Missing data points were then dealt with using statistical interpolation methods which were then normalized to provide consistency among the machine learning inputs of different scales.

Table 2. System Architecture Components

Layer	Sub-Modules	Functionality
Presentation Layer	Web UI, Dashboard	User interaction & visualization
Application layer	Prediction engine, Alert engine	Business logic & Analy cs
Data layer	Database, API Integration	Data storage & retrieval
External Services	Weather APIs	Real-time meteorological input

5. RESULTS:

The system integrates LSTM forecasting, an interactive dashboard, and alerts. LSTM achieved the best performance (91.6% accuracy, RMSE 1.2) on 52,560 observations, while the REST backend enabled efficient data access.

Table 3. Performance Evaluation of Prediction Models

Model	Accuracy (%)	RMSE	Predic on Type
ARIMA	82.5	2.1	Temperature
Random Forest	87.3	1.8	Weather Classifica on
LSTM	91.6	1.2	Mul variate Forecast



Fig.5.1. Historical & Predicted weather Trends



Fig.5.2. Modern Weather Intelligence UI

6. DISCUSSION

Using 52,560 hourly observations with a 70/15/15 split, LSTM achieved the best performance (91.6% accuracy, RMSE 1.2), followed by Random Forest (87.3%, RMSE 1.8) and ARIMA (82.5%). The results demonstrate the value of integrated forecasting and alerts, while limitations include data availability, limited spatial coverage, and lack of long-term forecasting.

7. CONCLUSION

The proposed Flask/FastAPI and React-based platform integrates weather data, ML forecasting, and real-time alerts. LSTM achieved 91.6% accuracy and 1.2 RMSE, while extreme events were detected in under 3 seconds. Future work will focus on IoT-based live data, cloud databases, long-term forecasting, and XAI.

8. REFERENCES

[1] M. Brown, "Weather portals and digital climate systems," *Journal of Atmospheric Informatics*, vol. 15, no. 2, pp. 45–53, 2021.

[2] A. Smith and R. Wilson, "Centralized climate data platforms," *International Journal of Climate Computing*, vol. 8, no. 1, pp. 10– 19, 2020.

[3] Sharma, U., Das, E., Diksha, & Yadav, P. (2025). *Interactive weather forecasting system using OpenWeather API and web technologies*. International Journal of Research – GRANTHAALAYAH, 13(1).https://doi.org/10.29121/granthaalayah.v13.i1.2025.6119

[4] R. Gupta and P. Singh, "Machine learning for temperature forecasting," *IEEE Access*, vol. 9, pp. 12345–12355, 2021, doi: 10.1109/ACCESS.2021.3056789.

[5] K. Lee, "Use of CSV-based climate datasets for environmental analysis," *Environmental Modelling & Software*, vol. 117, pp. 24– 32, 2019, doi: 10.1016/j.envsoft.2019.03.012.

[6] G. Kaur, V. K. Shukla, P. Verma, A. Verma, and C. T. S. Kumar, "Statutory Promise and Classroom Reality: A PRISMA Guided Systematic Review of the Legal Framework for Inclusive Education of Children with Disabilities in India," International Journal of Special Education, vol. 41, no. 14s, pp. 222–230, 2026.

[7] J. Liu, "Optimizing REST APIs for large environmental datasets," *IEEE Systems Journal*, vol. 16, no. 4, pp. 5890–5901, 2022, doi: 10.1109/JSYST.2021.3124567.

[8] H. Chen and D. Li, "Visualization Techniques for Climate and Weather Data Analysis," *IEEE Computer Graphics and Applications*, vol. 40, no. 6, pp. 72–81, 2020, doi: 10.1109/MCG.2020.3011456.

[9] P. Mishra and N. Verma, "REST-Based Backend Architectures for Meteorological Applications," *Int. J. Web Services Research*, vol. 17, no. 2, pp. 33–46, 2020.

[10] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2018.

[11] A. Verma, P. Verma, N. Mansoura, and S. Rajan, "Hybrid deep-learning model for resilient tourism enterprises," Annals of Tourism Research Empirical Insights, vol. 7, no. 1, p. 100220, 2026..

[12] T. Anderson, "Big Data Analytics for Climate Monitoring Systems," *IEEE Big Data Conference*, pp. 1345–1352, 2019.

[13] Zhu, X., Xiong, Y., Wu, M., Nie, G., Zhang, B., & Yang, Z. (2023). *Weather2K: A multivariate spatio-temporal benchmark dataset for meteorological forecasting*. arXiv preprint arXiv:2302.10493.

[14] L. Wang and Y. Zhao, "Real-Time Weather Alert and Forecasting Systems," *IEEE Access*, vol. 8, pp. 8632–8640, 2020, doi: 10.1109/ACCESS.2020.2987654.

[15] F. Perez and B. Granger, "Data Analysis and Processing with Pandas," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 13–21, 2019.

[16] A. Roy, "Backend API Design for Data-Driven Weather Applications," *International Journal of Computer Applications*, vol. 182, no. 15, pp. 22–26, 2020.

[17] S. Patil and R. Deshmukh, "Secure REST API Design for Environmental Data Systems," *IEEE Int. Conf. on Software Engineering*, pp. 155–160, 2022.

[18] Fu, D., Zhu, Y., Liu, Z., Zheng, L., Lin, X., & He, J. (2025). *ClimateBench-M: A multi-modal climate data benchmark*. arXiv preprint arXiv:2504.07394.

[19] M. Verma and P. Kulkarni, "Scalable Architectures for Climate Data Portals," *IEEE Software*, vol. 38, no. 5, pp. 44–50, 2021, doi: 10.1109/MS.2021.3079123.

[20] S. Kumar and A. Patel, "Cloud-Based Weather Data Storage and Analysis," *IEEE Int. Conf. on Cloud Computing*, pp. 210–215, 2020.

Biographies



Omkar Raskar completed his HSC from Ideal Junior College, Maharashtra. He is currently pursuing B.Tech. in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra



Prof. Sayali Belhe is an Assistant Professor in AI & Data Science at Vishwakarma Institute of Technology, Pune, and is pursuing her Ph.D. She holds B.E. and M.E. degrees in Computer Science and Information Technology.



Prathamesh Kumbhare completed his HSC from Kendriya Vidyalaya Ganeshkhind, Pune, Maharashtra. He is currently pursuing a B.Tech. in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra



Sandesh Rabade completed his HSC from Jai Bhavani High School, Molkheda, Maharashtra. He is currently pursuing a B.Tech. in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra



Shaikh Sameer Tabrez completed HSC from Alpha Jr. College of Science and Commerce, Mumbai, Maharashtra. He is currently pursuing Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra



Hitesh Pawar completed his HSC from K V Pendharkar, Dombivali, Maharashtra. He is currently pursuing B.Tech. in Artificial Intelligence and Data Science from Vishwakarma Institute of Technology, Pune, Maharashtra

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Review of Lightweight and Explainable AI for Clinical Edge Deployment

Shital Ajit Dumbre[1], Dr. Arati Kale[2]

[1],[2]School of Computing, MIT Art, Design and Technology University, Loni, Maharashtra, India, shital.dumbre@mituniversity.edu.in, arati.kale@mituniversity.edu.in

Abstract.

Clinical artificial intelligence models are highly accurate in their diagnostic outcomes, but have difficulty in their practical application because of the high computing resources and non-interpretable nature. The paper will solve this issue of empowering effective and reliable AI towards clinical edge conditions using limited resources.

To address this, a systematic comparative study of lightweight AI methods such as pruning, quantization, knowledge distillation and sparse architectures is discussed and explainable AI approaches such as SHAP, LIME and Grad-Cam. Performance, efficiency, and interpretability are evaluated using a single evaluation framework.

It is analyzed that lightweight models can reach a reduction of up to 80-95 percent in model size and enormous latency cuts and still ensure a high level of performance within a narrow performance error. These results justify the implementation of AI models to edge images like microcontrollers, smartphones, and embedded health care frameworks to the application of diagnostics and remote monitoring and clinical decision support.

Keywords. Clinical AI; Model Compression; Knowledge distillation; explainable AI; Edge computing;

1. INTRODUCTION

Deep learning has enhanced medical diagnosis, particularly in imaging and patient monitoring, but application in clinical practice is a challenge. Models with high capacity that are trained to be as accurate as possible may need to compute on the cloud and cannot meet the on-device, real-time, and privacy requirements of regulations like GDPR and HIPAA. This has seen a shift towards deployment-first AI, where efficiency, energy consumption, and compatibility with various edge devices are critical [1]. Meanwhile, the black-box characteristics of deep models prevent clinical trust, as clinicians require clear and comprehensible explanations to make serious judgments. These issues have made lightweight and explainable AI models more popular, which can be implemented on edge devices, including wearables, handheld scanners, and smartphones. This survey evaluates compact model design, compression and sparsity, explainability, real-world healthcare applications and open challenges in 2021-2025.

2. RELATED WORKS

2.1. Compression Methods

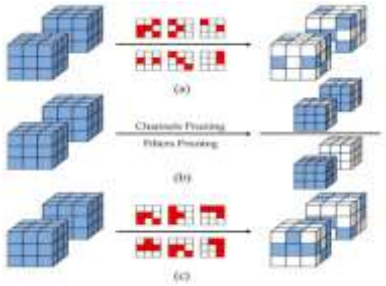


Fig. 2.1.1. CNN pruning techniques: non-structured, structured, and pattern pruning.

Pruning, quantization, and knowledge distillation are the key methods that are required to compress models and deploy AI to edge devices. These techniques can greatly cut down the size of models, but do not impact performance: pruning removes unnecessary weights to build sparse networks, quantization transfers values of high precision to lower-bit representations to save memory, and distillation transfers knowledge in large teacher networks to smaller student networks [4]. Recent healthcare research reports a good performance, with models scaled down to 1518 percent of their initial size and have up to 56 inference times faster with little to no loss of accuracy [5], [6]. Practically, the best performance is achieved by merging the techniques (e.g., pruning with quantization) and allows reducing the number of parameters by up to 90 percent with only a slight decrease in performance, allowing edge deployment into healthcare environments to be a practicable reality [9].

2.2. Sparse Architectures and Small Models with XAI

Lightweight model design provides an alternative to conventional compression by constructing efficient architectures based on fundamentals. Factorized convolutions and scaling in models like MobileNet, EfficientNet, and ShuffleNet can reduce the number of parameters and computation by a large margin compared to heavy networks like VGG-16 [10]. Recent health signal architectures such as MoE-TinyMed show that sparse architectures can be more efficient than larger dense models with fewer active parameters, emphasizing the power of conditional computation [12]. Explainable methods (e.g., Grad-CAM) can still be used even with compressed models, which can be trusted by clinicians. The real-world applications indicate an increasing possibility of edge-based clinical AI, as TinyML systems can run on low power (around 100 mW) and the smartphones can infer in 50-100 ms. Nevertheless, the literature does not provide standard evaluation frameworks that can be used to collectively evaluate baseline models, compression techniques, and explainability, which is an open research gap.

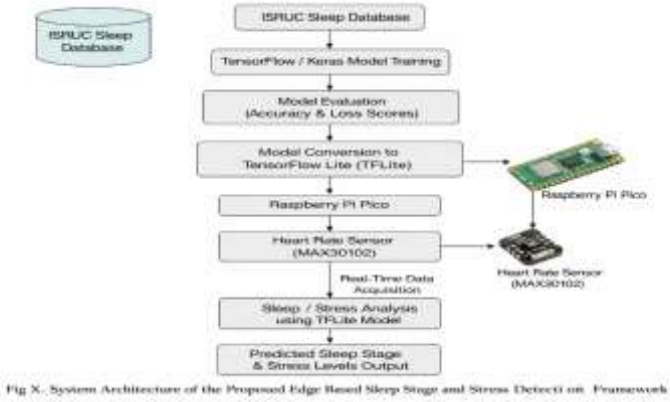


Fig. 2.2.1. TinyML pipeline including data collection, training, quantization, and deployment.

3. METHOD, EXPERIMENTS AND RESULTS

3.1. Methodology

The method includes three main steps: (1) literature search and classification, (2) comparative performance of the model optimization methods, and (3) assessment according to metrics related to deployment. To begin with, the systematic review included recent works published between 2021-2025, centering on the use of AI in the clinical imaging field, natural language processing, and physiological signal analysis. Second, a comparative framework was created to compare these approaches on their capacity to decrease the computational complexity and retain the diagnostic performance. Lastly, the techniques have been assessed on the basis of a common set of metrics representing practical deployment conditions, namely accuracy, model size, latency, energy usage and interpretability.

3.2. Evaluation Metrics

An effective assessment framework is needed to evaluate lightweight clinical AI systems. In the paper, several performance dimensions are taken into account:
Accuracy / F1-score / AUC: Diagnostic and predictive performance.
Model Size (MB): Number of storage requirements of the edge deployment.
Latency (ms): Inference time, which is important in real-time applications.
Energy Consumption (mW): Of significance to wearable and battery-powered devices.
Compression Ratio: Level of downsizing of model over baseline.
Explainability Metrics: Incorporating IAUC, DAUC, and saliency consistency.

3.3. Results and Observations

The comparative analysis shows that there are a few significant trends. First, compression methods like pruning and quantization of models are highly effective in minimizing the computational cost at the expense of reasonable accuracy. Losss of up to 80-95% in model size with little performance loss is obtained in most cases. Second, the lightweight and sparse architecture is an alternative to using post-training compression, which includes efficiency into model design. MobileNet and EfficientNet architectures can be highly accurate with fewer parameters, whereas mixture-of-experts models can do conditional computation, achieving greater efficiency without compromising accuracy. Third, the methods of

explainability still work even when compressed models are used. Lastly, deployment-related studies show that lightweight models can be deployed to edge devices, including microcontrollers, smartphones, and embedded systems, and can provide real-time inference with a low power budget.

4. DISCUSSIONS

Although the development is fast, there are several issues that are not addressed:

- **Integration and Regulatory Compliance:** One of the gaps in the reviewed literature is the lack of standardized comparative reporting. A lot of research introduces compressed or lightweight models without a consistent comparison to strong baseline architectures or recent state-of-the-art algorithms as well as popular evaluation protocols. Subsequent work ought to put more focus on reproducible benchmark configurations, which collectively report on predictive, computational, and interpretability results. The connection between compressed models and hospital systems (PACS, EHR) has not been greatly discussed. This can be filled in the future, e.g. with designing interoperable pipelines and the requirements of FDA/CE. Improved XAI standards and user research are required clinical adoption requirements to support physicians in their trust and proper interpretation of model results.
- **Energy and Runtime Benchmarking:** It is not common to have overall benchmarks of a wide range of hardware. Community intervention is required to make comparisons between models on standard medical problems (e.g. chest X-ray classification, ECG analysis) on devices (microcontrollers, RPi, phones) based on latency, energy and memory. This would inform architecture/hardware co-design.

5. CONCLUSIONS

1. **Problem Statement / Motivation:** High computational demands and limited hardware resources and interpretability are the reasons why clinical AI models are not easily deployed.
2. **Method Used:** A systematic comparative study of lightweight AI methods such as pruning, quantization, knowledge distillation, sparse architectures and explainable AI methods was performed.
3. **Key Findings:** Lightweight models offer major advantages in reducing model size (80-95%), latency and competitive performance. The efficiency of sparse architecture is even more efficient and surpasses large architecture in some situations. Explainability methods enhance transparency and aid clinical adoption.
4. **Limitations and Future Work:** The study is also disadvantaged by the absence of standardized benchmarks and similar evaluation techniques. The future research efforts should be on standardized benchmarking systems, incorporation of privacy-preserving methods and clinical feasibility.

6. REFERENCES

[1] C. M. Nguyen and T.-S. Hy, “Efficient Deep Learning for Medical Imaging: Bridging the Gap Between Research Environments and Clinical Deployment,” arXiv preprint arXiv:2602.00910, 2026.

[2] E. Paula et al., “Comparative analysis of model compression techniques for achieving carbon efficient AI,” Sci. Rep., vol. 15, p. 23461, 2025.
[3] C. Johnvictor et al., “TinyML-Based Lightweight AI Healthcare Mobile Chatbot Deployment,” J. Multidiscip. Healthc., vol. 17, pp. 5091–5104, 2024.
[4] P. Srivastava, N. Shah, and K. Jaiswal, “Microcontroller-Based EdgeML: Health Monitoring for Stress and Sleep via HRV,” Engineering Proceedings, vol. 78, no. 1, 2024, art. 3.
[5] E. F. Rivera-Guzmán et al., “Quantization of Deep Neural Networks for Medical Image Analysis: A Systematic Review and Meta-Analysis,” Technologies, vol. 14, no. 1, art. 76, 2026.
[6] F. M. Alabbasy et al., “Compressing medical deep neural network models for edge devices using knowledge distillation,” J. King Saud Univ. – Comput. Inf. Sci., vol. 35, no. 7, 2023, art. 101616.
[7] X. Sui et al., “A Hardware-Friendly High-Precision CNN Pruning Method and Its FPGA Implementation,” Sensors, vol. 23, no. 2, art. 824, 2023.
[8] N. Malik, “Interpretability-Aware Pruning for Efficient Medical Image Analysis,” arXiv, 2025 (preprint).
[9] M. S. Vani et al., “Personalized health monitoring using explainable AI: bridging trust in predictive healthcare,” Sci. Rep., vol. 15, p. 31892, 2025.
[10] T. Dutta et al., “Towards On-Device Explainable Artificial Intelligence for the Semantic Web of Everything,” Future Gener. Comp. Syst., 2026 (preprint).
[11] D. Shehada et al., “An Explainable Framework for Mental Health Monitoring Using Lightweight and Privacy-Preserving Federated Facial Emotion Recognition,” Sensors, vol. 25, no. 7, 2024.
[12] S. Jiang et al., “MoE-TinyMed: Mixture-of-Experts for Tiny Medical Large Vision-Language Models,” arXiv preprint arXiv:2404.10237, 2024.

Biographies



Shital Ajit Dumbre is an Assistant Professor in the Department of Computer Science and Engineering at MIT Art, Design and Technology University, Pune, India. She holds a M.E. in Computer Science and Engineering Her research interests include Artificial Intelligence, Machine Learning, Data Science, Healthcare Analytics, and Explainable AI.



Arti Kale is an Assistant Professor in the Department of Computer Science and Engineering at MIT Art, Design and Technology University, Pune, India. She holds a Ph.D. in Computer Science and Engineering Her research interests include Artificial Intelligence, Machine Learning, Data Science, Healthcare Analytics, and Explainable AI.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Hybrid Crowdsourced Platform for Civic Problem Reporting and Resolution

Tejaswini Bhosale[1], Vishal Hingmire[2], Rajita Dixit[3], Vaidehi Nadge[4],Vaishnavi Jadhav[5],Jayashree Rajesh Prasad[6]

[1,4,5,6] School of Computing, MIT Art Design and Technology, Pune , tejaswini.bhosale@mituniversity.edu.in [1], vaidehinadge185@gmail.com [4], vaishnavi1352005@gmail.com[5], jayashree.prasad@mituniversity.edu.in[6]

[2]Department of Electronics and Telecommunication Engg, Arvind Gavali College of engineering, Satara, Vs.hingmire@gmail.com [2]

[3]Centre for Distance and Online Education, Bharati Vidyapeeth Deemed to be University, Pune , Rajita.Dixit@bharatividyapeeth.edu [3]

Abstract

Urban civic management often suffers from slow reporting, limited transparency, and weak coordination between citizens and authorities due to manual complaint processes. This study proposes a real-time, crowdsourced civic issue reporting and resolution system that enables citizens to report road damage, littering, water leaks, and infrastructure failures through mobile or web platforms. Authorities can assign and monitor complaints using dashboards, while public maps, notifications, and reward mechanisms improve transparency and participation. The system supports faster resolution, stronger accountability, and evidence-based civic management.

Keywords. Crowdsourcing, Civic Engagement, Conflict Reporting, Civil. Society, Civic Issue Reporting.

1 INTRODUCTION

Rapid urbanization and inadequate infrastructure create civic issues such as waste, water wastage, damaged roads, traffic accidents, and broken streetlights, affecting quality of life and public trust. Mobile-based crowdsourcing enables timely reporting and monitoring while improving citizen-authority communication. However, existing e-governance systems often face multilingual, low-literacy, verification, duplication, and resolution-tracking challenges. The proposed system addresses these gaps through structured reporting, geolocation validation, decentralized verification, smart prioritization, and an open dashboard, enabling transparent resolution tracking, better decision-making, and active citizen participation.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2 BACKGROUND STUDY

Previous studies have examined cybersecurity coordination [1], civic engagement and crowdsourced journalism [2], community resilience and trust [3], citizen reporting through digital channels [4], customized web technologies [5], and crowd-civic systems [6]. Other research has addressed ethnic and communal conflicts in Assam [7], Hindu-Muslim conflict in India [8], citizen media for documenting human rights abuses [9], and insurgency in Northeast India [10]. However, existing studies provide limited integration of real-time reporting, verification, classification, prioritization, allocation, and tracking. The proposed system addresses this gap through a citizen-friendly platform with an issue-prioritization algorithm based on urgency, frequency, resolution time, and impact, supporting transparency, timely response, citizen engagement, and scalable smart-city governance.

3 METHODOLOGY

The proposed system of Crowd-Sourced Civics Issues Reporting and Resolution System abides by scalable and modular architecture that permits real-time reporting of issues, validation, and monitoring. Citizen Interface (Mobile/Web Application) enables citizens to share issues along with descriptions, photos and comments.

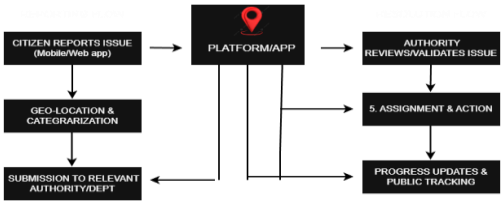


Figure 3.1. Proposed hybrid crowdsourcing system architecture.

After an issue is reported, GPS automatically records its latitude and longitude. Each issue is represented as $i_k = (L_k, C_k, U_k, F_k, T_k)$, with location, category, urgency, frequency, and reporting time, respectively. The platform stores reports, removes duplicates, prioritizes issues, and supports citizen-government communication. The processing engine performs geolocation tagging and categorization, while the severity score is calculated as:

$$S_k = \alpha U_k + \beta F_k + \gamma T_k + \delta I_k \quad (1)$$

where (I_k) represents the issue impact and the coefficients determine factor weights.

Authorities review and validate reports through the dashboard. Valid issues are approved and automatically assigned to the appropriate department based on category and location:

$$D(i_k) = \int (C_k, L_K) \quad (2)$$

Progress is monitored through resolution time $R_k = T_{close,k} - T_{report,k}$. The mean resolution time is calculated across all reported issues to evaluate overall response efficiency. Public tracking

provides transparency by allowing citizens to monitor the status and resolution of submitted issues.

$$R_{avg} = \frac{1}{N} \sum_{K=1}^{N} R_k \quad (3)$$

The system was developed on an Intel Core i5 computer with 8 GB RAM, 512 GB SSD, and Windows 11, using Python with the Flask framework. TensorFlow models and the Google Maps API were used for AI-based processing and GPS functionality. The dataset was sourced from World Bank [11] and the Ministry of Housing and Urban Affairs, Government of India [12].

4.3 Result Analysis & Discussion

The proposed system achieved over 92% accuracy in issue classification and prioritization, enabling correct categorization and timely routing to relevant authorities. Garbage complaints were the most frequent, followed by electricity, road, and water issues, highlighting key areas for intervention. Most complaints received medium priority, while urgent public-safety issues were assigned higher priority for efficient resource allocation. The system supports faster resolution, improved governance, and citizen participation. Future integration with advanced technologies and civic platforms could further improve performance and scalability.

Table 2. Issue Classification Accuracy

Issue Category	Total issues	Correctly Classified	Accuracy(%)
Road	**120**	**112**	**93.3**
Garbage	150	142	94.6
Water	**100**	**92**	**92.0**
Electricity	130	121	93.0

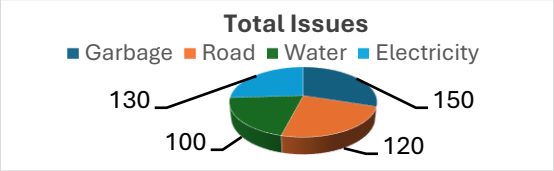


Figure 4.2. Distribution of Reported Civics Issues by category.

Table 3. Priority Level Distribution

Priority Level	**Number od Issues**
High	180
Medium	220
Low	100

4 CONCLUSIONS

The proposed hybrid civic crowdsourcing system enables real-time reporting, classification, and resolution of urban civic problems, addressing delays, poor prioritization, and limited

transparency in traditional complaint systems. Issues are prioritized using urgency, frequency, time, and impact to support structured decision-making. Citizen participation, automated assessment, and real-time status updates improve operational efficiency, transparency, and public trust. Overall, the system transforms conventional complaint management into a data-driven civic intelligence platform, with future potential for predictive analytics and integration with smart-city technologies.

5 REFERENCE

[1] Bahuguna, Ashutosh. "FIRe: Firefox for Computer Security Incident Reporting and Coordination." IITM J. Manage. IT 6 (2015): 3-11.

[2] Mitchell, Scott SD, and Merlyna Lim. "Too crowded for crowdsourced journalism: Reddit, portability, and citizen participation in the Syrian crisis." Canadian Journal of Communication 43.3 (2018): 399-419.

[3] Braun-Lewensohn, Orna, Sarah Abu-Kaf, and Tehila Kalagy. "What factors explain anger and mental health during the COVID-19 pandemic? The case of Israeli society." World journal of psychiatry 11.10 (2021): 864.

[4] Mergel, Ines. "Distributed democracy: SeeClickFix. com for crowdsourced issue reporting. 2012." (2012): 27.

[5] Vreÿ, Francois. "Maritime aspects of illegal oil-bunkering in the Niger Delta." Australian Journal of Maritime & Ocean Affairs 4.4 (2012): 109-115.

[6] McInnis, Brian, et al. "Crowdsourcing law and policy: a design-thinking approach to crowd-civic systems." Companion of the 2017 ACM conference on computer supported cooperative work and social computing. 2017.

[7] Chakraborty, Joya, Anjuman Borah, and Muktikam Hazarika. "FROM FRACTURES TO FRAMES: CONFLICT REPORTING IN NEWSPAPERS OF ASSAM." Global Media Journal: Indian Edition 6 (2015).

[8] Singh, Arshdeep. "Hindu-Muslim Conflict in India: Causes, Consequences & Resolution." International Journal of Social Science Research & Review 7.10 (2024): 62-79.

[9] Williams, Erica Frances. Using Citizen Media and Open Source Investigations to Promote Human Rights. Diss. UC Berkeley, 2017.

[10] Sahni, Ajai. "Survey of conflicts & resolution in India's Northeast." FAULTLINES-NEW DELHI- 12 (2002): 39-112.

[11] World Bank, Citizen Engagement and Feedback Systems in Urban Governance, Washington, DC, USA, 2016.

[12] Ministry of Housing and Urban Affairs, Smart Cities Mission Statement and Guidelines, Government of India, 2015.

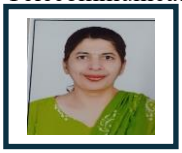
Biographies



Dr. Tejaswini Bhosale is an Assistant Professor at MIT Art, Design and Technology University, Pune, and a completed her PhD from Birla Institute of Technology, Mesra.



Dr. Vishal Sharad Hingmire is an Associate Professor in the Department of Electronics & Telecommunication Engineering at Arvind Gavali College of Engineering, Satara, Maharashtra.



Dr. Rajita Dixit is an Assistant Professor, Bharati Vidyapeeth, Pune, India. Her academic and research interests encompass management, digital education, e-learning & AI in creative and organizational domains.



Vaidehi Nadge is a student at the School of Computing, MIT Art, Design and Technology University, Pune.



Vaishnavi Jadhav is a student at the School of Computing, MIT Art, Design and Technology University, Pune.



Dr. Jayashree Rajesh Prasad is Associate Dean–Research and Head of the Department of Computer Science & Engineering, School of Computing, MIT ADT University, Pune.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

AI-Driven Vision-Guided Robotic Arm for Real-Time Defect Detection in Smart Manufacturing

Mrs. Aparna Prakash Bairagi, Dr. S. K. Sonkar

Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner, MH, India, aparnasbairagi03@gmail.com, sonkar83@gmail.com

Abstract.

Manual quality inspection is time-consuming and may miss small defects due to worker fatigue. This study proposes an automated inspection and sorting system using a camera, YOLO, CNN, CFNN, and robotic arm. The camera captures product images, YOLO detects object locations, and CNN identifies defects. CFNN classifies products as acceptable or defective, after which the robotic arm sorts them accordingly. Testing showed faster and more reliable inspection than manual methods, reducing human error, improving sorting efficiency, and supporting consistent product quality.

Keywords. Defect Detection, Robotic Arm, Computer Vision, Industrial Automation, Smart Manufacturing.

1. INTRODUCTION

Industry 4.0 has transformed manufacturing through automation, artificial intelligence (AI), robotics, connected sensors, and data-driven decision-making. Automated quality inspection is increasingly important because manual inspection is time-consuming, labor-intensive, and susceptible to fatigue and human error. Small defects such as cracks, scratches, dents, and surface marks may therefore remain undetected, affecting product quality and customer satisfaction [1-5]. Computer vision and deep learning provide effective solutions by enabling automated detection and classification of manufacturing defects. YOLO can locate products in real time, while Convolutional Neural Networks (CNNs) extract visual features for defect identification. A Convolutional Fuzzy Neural Network (CFNN) can further improve classification by handling complex defect patterns. However, integrating vision-based detection, deep learning, and robotic sorting into a unified system remains challenging [6-10]. This study proposes an AI-powered robotic inspection system that combines YOLO, CNN, and CFNN with a robotic arm for automated defect detection and sorting, aiming to improve inspection accuracy, reduce manual effort, and enhance production efficiency [11-14].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

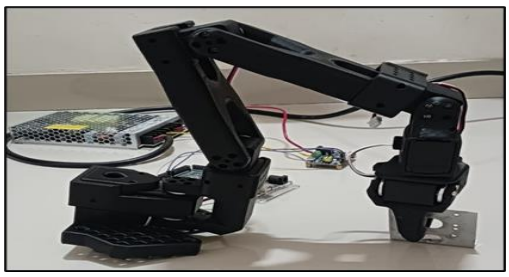


Fig.1.1 Structure of Robotic Arm

2. RELATED WORK

Recent studies have demonstrated the growing application of deep learning, computer vision, and robotics for automated manufacturing inspection. In 2025, a study proposed an improved VGG16 model for detecting manufacturing defects such as cracks and scratches. The model showed improved performance under varying lighting conditions and outperformed traditional inspection approaches [1]. Lin et al. (2024) developed a vision-based robotic arm for real-time defect detection and object classification using YOLO. Their system successfully detected and sorted products on production lines, demonstrating its potential for industrial applications [2]. Keshvarparast et al. (2023) reviewed collaborative robots in manufacturing and assembly, highlighting their benefits in safety, flexibility, and human-robot collaboration, while emphasizing the need for advanced sensing and learning capabilities [3]. Santos et al. (2024) developed a camera-guided robotic pick-and-place system capable of accurately locating and handling objects despite positional variations [4]. Similarly, Kim and Park (2023) proposed a vision-based inspection system that automatically detects surface defects during manufacturing [5]. These studies demonstrate significant progress in automated inspection and robotic systems; however, integrated systems combining YOLO-based detection, CNN-based feature extraction, CFNN classification, and automated robotic sorting remain limited.

3. METHOD, EXPERIMENTS AND RESULTS

3.1. Proposed Method

The proposed system automatically inspects products on a production line and separates defective products using computer vision and a robotic arm. An Arducam 8MP camera captures product images, which are preprocessed through noise removal, enhancement, and resizing. YOLO detects and localizes the product, after which a CNN extracts visual features and identifies defects such as cracks, scratches, dents, and damaged areas. The extracted features are classified by a Convolutional Fuzzy Neural Network (CFNN) as OK (normal) or NOK (defective). Based on the classification result, the robotic arm automatically sorts acceptable and defective products. Figure 1 illustrates the complete inspection and sorting workflow.

The experimental setup included an Arducam 8MP camera, Arduino Mega 2560, serial bus driver board, servo-motor robotic arm, and GPU-enabled laptop. Python, OpenCV, YOLO,

CNN, CFNN, and Arduino IDE were used for system development and processing. The dataset contained industrial component images categorized into OK and NOK classes and was divided into training and validation sets for model development and performance evaluation.

3.2. Experimental Results

The experiments demonstrated effective automatic quality inspection and classification. Training and validation accuracy increased gradually across epochs, indicating successful learning of defect patterns with minimal overfitting. Figure 3.2.1 presents the accuracy trends of the proposed system. The dataset analysis further showed that products were correctly classified into OK and NOK categories. As shown in Figure, most samples belonged to the OK class, while fewer were classified as NOK, confirming the system's ability to distinguish acceptable and defective products.

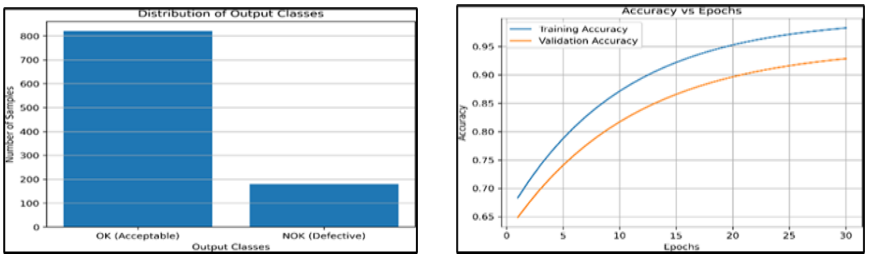


Fig. 3.2.1. Accuracy vs. epochs, Distribution of Output Classes

The combined YOLO, CNN, and CFNN approach improved defect detection and automated sorting. YOLO localized products, CNN identified surface defects, and CFNN classified them as acceptable or defective. The robotic arm then sorted products accordingly. The proposed system reduces manual inspection effort and human error while improving inspection accuracy and production efficiency. Integrating AI-based vision with robotic automation supports reliable quality control and demonstrates the practical potential of Industry 4.0.

4. CONCLUSION

Manual defect inspection really drags down the speed of manufacturing and, let's be honest, people make mistakes. This project aims to change that by building an automated quality control system that's actually smart. So, what's under the hood? It's a blend of computer vision, deep learning, and robotics all working together. The system uses YOLO to spot objects, CNNs to pull out important details, and CFNNs to figure out what kind of defect it's looking at.

5. REFERENCES

[1] A. Baiganova, Z. Ubayeva, Z. Taskalyeva, and S. B. Nurgaliyev, ""Automated Defect Detection in Manufacturing Utilizing Enhanced VGG16 Convolutional Neural Networks," International Journal of Advanced Computer Science and Applications, vol. 16, no. 1, pp. 102–110, 2025.

[2] C. J. Lin, J. Y. Jhang, Y. J. Gao, and H. MHuang, "Vision-Based Robotic Arm in Defect Detection and Object Classification Applications," Sensors and Materials, vol. 36, no. 2, pp. 421–436, 2024.
[3] A. Keshvarparast, D. Battini, O. Battaia, and A. Pirayesh, “Collaborative Robots in Manufacturing and Assembly Systems: Literature Review and Future Research Agenda,” *Journal of Intelligent Manufacturing*, vol. 35, pp. 2065–2118, 2024.
[4] A. A. Santos, C. Schreurs, and A. Ferreira da Silva, “"Combining Artificial Vision and Image Processing into a Pick-and-Place Collaborative Robotic System," Journal of Intelligent Robotic Systems, vol. 110, no. 2, pp. 1–14, 2024.
[5] S. W. Kim and H. J. Park, “Machine Vision-Based Defect Detection in Smart Manufacturing Using Deep Learning,” *IEEE Access*, vol. 11, pp. 28741–28752, 2023.
[6] A. Gupta and S. Patel, “Computer Vision-Based Automated Defect Detection System for Industrial Applications,” *IEEE Access*, vol. 12, pp. 21011–21022, 2024.
[7] V. D. Cong, L. D. Hanh, and L. Phuong, “Design and Development of Robot Arm System for Classification and Sorting Using Machine Vision,” *FME Transactions*, vol. 50, no. 3, pp. 455–463, 2022.
[8] S. Zhang, Y. Chen, and H. Li, “Smart Manufacturing Inspection Using Artificial Intelligence and Robotics,” *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 211–220, 2024.
[9] Y. Liu, J. Zhang, and Z. Li, “Real-Time Industrial Surface Defect Detection Using Deep Learning Networks,” *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1285–1294, 2023.
[10] S. Chen and Y. Liu, “AI-Based Visual Inspection System for Smart Manufacturing,” *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 4, pp. 2175–2186, 2023.
[11] P. Kumar and M. Verma, “AI-Based Industrial Defect Detection and Robotic Sorting System for Smart Factories,” *IEEE Access*, vol. 13, pp. 41200–41212, 2025.
[12] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, “Industry 4.0 Technologies and Their Applications in Manufacturing: A Review,” *IEEE Access*, vol. 10, pp. 119602–119620, 2022.
[13] G. Aceto, V. Persico, and A. Pescapé, “Industry 4.0 and Smart Manufacturing: A Review of Research Issues and Applications,” *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3211–3224, 2022.
[14] V. T. Nguyen, P. T. Nguyen, and S. F. Su, “Vision-Based Pick-and-Place Control System for Industrial Robots Using an Eye-in-Hand Camera,” *IEEE Access*, vol. 13, pp. 12845–12858, 2025.

Biographies



Mrs. Aparna Prakash Bairagi received the bachelor's degree in computer engineering from Savitribai Phule Pune University in 2017, the master's degree in computer engineering from Savitribai Phule Pune University in 2026, respectively. She is currently working as an Assistant Professor at the Department of Computer Engineering, Faculty of Engineering, Dbatu University. She is research areas include AI-robotics, deep learning, and machine learning.



Dr. Shrinivas k. Sonkar received the bachelor's degree in computer engineering from SRTM university. in 2005, the master's degree in computer engineering from MGM's college of Engg. Nanded. University in 2010, received the PHD in computer engineering from Savitribai Phule Pune University. in 2020 respectively. He is currently working as an Associate Professor at the Department of Computer Engineering, Faculty of Engineering, SPPU University

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Explainable Longitudinal Machine Learning Framework for Early Cervical Cancer Risk Prediction

Sneha K. Deshmukh[1], Dr. Suvarna Eknath Pawar[2], Sachin Arun Thanekar[3], Komal Munde[4]

Department of Computer Science and Engineering,

MIT Art, Design and Technology University, Loni Kalbhor, Pune, India
[1]khupsesneha68@gmail.com, [2]suvarna.pawar@mituniversity.edu.in,
[3]sachin.thanekar@mituniversity.edu.in, [4]komal.munde@mituniversity.edu.in

Abstract

Early cervical cancer detection can reduce mortality by enabling timely intervention. This study develops an interpretable risk-prediction framework using SVM, Random Forest, XGBoost, and longitudinal LSTM models on clinical and behavioral data. Class imbalance is addressed through feature standardization and class-weighted learning. SHAP provides global and patient-specific explanations, with age, reproductive history, and contraceptive use identified as important risk factors. The results show strong ensemble performance, while LSTM effectively captures temporal risk patterns.

Keywords: Cervical Cancer, Early Risk Prediction, Explainable Artificial Intelligence, SHAP, Long Short-Term Memory

1. INTRODUCTION

Cervical cancer remains a major health concern, particularly in low-resource settings where screening and early diagnosis are limited [1]. Machine learning (ML) supports risk prediction using demographic, clinical, and behavioral data, while XAI techniques such as SHAP and LIME improve model transparency [4,6,8–10]. However, many existing methods overlook changes across clinical visits. This study integrates ML, LSTM-based temporal learning, threshold optimization, and SHAP for interpretable cervical cancer risk prediction.

2. RELATED WORK

Machine learning models such as Random Forest, Gradient Boosting, and SVM have shown promising results for cervical cancer risk prediction using clinical and behavioral data [1,8]. SMOTE and ADASYN improve minority-risk detection, while SHAP and XAI enhance prediction interpretability [1,8]. LSTM-based models capture longitudinal EHR information [12], while radiomics and multimodal approaches provide additional predictive capabilities [13]. However, missing data, preprocessing, external validation, and temporal modeling remain key challenges [11].

3. METHODOLOGY

Figure 1 shows the overall framework of the proposed explainable longitudinal framework, which involves preprocessing, baseline modeling, LSTM prediction, threshold optimization, and interpretability analysis

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International



Figure 3.1. Architecture diagram

3.1. Dataset Description and Preprocessing

This study relies on structured cervical cancer data comprising of demographic, behavioral and clinical attributes. Median and Mode imputation were used to deal with missing values. N-score was used to normalize numerical attributes.

$$X_{scaled}=\frac{X-\mu}{\sigma}$$

in which μ and σ are the mean and standard deviation, respectively. Class weighted learning was used to reduce the class imbalance and stratified sampling was used to ensure the data sets for training and testing were representative and had class proportions.

3.2. Baseline Machine Learning Models

The baseline machine learning models were compared to each other by cross validation, with RF, XGBoost, and SVM. They were evaluated with Accuracy, Recall, F1-score and ROC-AUC for a complete assessment.

3.3. Longitudinal LSTM Model

To capture the temporal risk progression, a Long Short-Term Memory (LSTM) system was needed to be developed that would serve as a Longitudinal LSTM Model A (LSTM). The data given as input were reorganized into a 3-D tensor

$$(\eta_{samples,}\ \eta_{timesteps,}\ \eta_{features})$$

3.4. Decision Threshold Optimization

Though the default probability threshold (0.5) is used, threshold tuning was carried out to maximise F1-score and Recall. The optimum threshold was found to be:

$$t^{*}=\arg\max_{t\in[0,1]} F1(t)$$

This guarantees a better sensitivity, which is essential in medical diagnosis to avoid false negativity.

3.5. Explainability Using SHAP

SHAP was applied to provide global and patient-level interpretation with the help of Summary plots and Force plots using TreeExplainer and DeepExplainer to make predictions more transparent and clinically understandable.

3.6. Evaluation Metrics

Model performance was assessed using F1-score, Accuracy, Precision, Recall, and ROC-AUC, measuring overall correctness, high-risk identification, sensitivity, balance, and class discrimination across thresholds.

4. RESULTS AND DISCUSSION

There is limited multicollinearity between some features as can be seen in Figure 2, as the moderate correlations between age and pregnancies (0.52) and age and first intercourse

(0.36) suggest. For the cervical cancer prediction model, Figure 3 presents the SHAP summary and the contributions of the model's features to the risk of cervical cancer. Age, first sexual intercourse, sexual partners and contraceptive duration are the most important predictors, and smoking is comparatively less important.

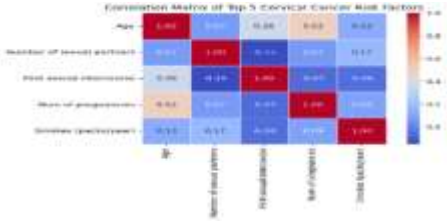


Figure 4.1: Correlation Matrix of Cervical Cancer Risk Factors

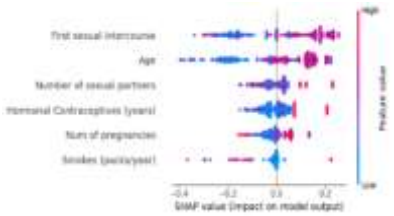


Figure 4.2: SHAP summary plot

Mean absolute SHAP values are displayed in Figure 4 by the features. Influence is strong for age at first intercourse and age, moderate for sexual partners and contraceptive use. In addition to this, pregnancy and smoking add to this. Temporal risk progression is illustrated in Figure 5, as patient 1 shows an increasing trend over the 4 visits and patient 2 consistently low risk.

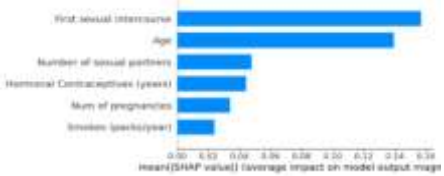


Figure 4.3: Importance of Global Features by the Mean Absolute SHAP Values to predict the risk of cervical cancer.



Figure 4.4: Patient risk Trajectories

TABLE I. COMPARISON OF PERFORMANCE

Model	Precision	Recall	F1-score	Accuracy	AUC
SVM	0.81	0.74	0.77	0.89	0.88
Random Forest	0.85	0.79	0.82	0.92	0.91
XGBoost	0.87	0.81	0.84	0.93	0.93
LSTM	0.80	0.89	0.84	0.90	0.92

At the 0.5 threshold, LSTM has a high accuracy of 0.90 and a high recall of 0.84, as shown in Table 1. With optimization, this recalled 0.89, enhancing sensitivity to early risk detection. It has good predictive ability with 0.92 ROC-AUC and 0.84 F1-score.

5. CONCLUSION

This study introduces a longitudinal match learning approach to the early prediction of cervical cancer risk based on clinical and behavioral features. Several machine learning models were tested (SVM, Random Forest, XGBoost, LSTM) and the SHAP model gave

interpretable predictions. The results show good ensemble results and LSTM can well reflect temporal risk patterns and enhance the sensitivity of early screening.

7. REFERENCES

[1] R. Shakil et al., “A precise machine learning model: Detecting cervical cancer using feature selection and explainable AI,” J. Pathol. Inform., vol. 15, Dec. 2024.
[2] P. Karthikeyan et al., “Explainable AI based cervical cancer prediction using FSAE feature engineering and H2O AutoML,” Sci. Rep., vol. 15, 40439, 2025.
[3] S. Prusty et al., “Predicting cervical cancer risk probabilities using advanced H2O AutoML,” PeerJ Comput. Sci., vol. 10, e1916, 2024.
[4] W. Chansanam et al., “An analytics-based framework for early detection of cervical cancer,” Sci. Dir. (2025).
[5] “An ensemble classification approach for cervical cancer prediction using behavioral risk factors,” Healthc. Anal., vol. 5, 100324, 2024.
[6] R. Ashtagi et al., “Cervical Cancer Prediction Using Machine Learning,” J. Electr. Syst., vol. 20, No. 1s, 2024.
[7] S. Abrar et al., “Recent advances in applications of machine learning in cervical cancer research,” Obstet. Gynecol. Sci., 2025.
[8] P. Roy et al., “Interpretable artificial intelligence (AI) for cervical cancer risk analysis leveraging stacking ensemble and expert knowledge,” Digit. Health, vol. 11, 2025.
[9] K. Toffaha et al., “Enhancing cervical cancer risk prediction using machine-learning and explainable models,” J. Med. Data Health, 2025.
[10] J. Su et al., “Cervical cancer prediction using routine blood analysis and SHAP interpretation,” 2024.
[11] Y. Wu and A. Wang, “A review on the applications of artificial intelligence for diagnosing and treating cervical cancer,” Int. J. Womens Health, 2025.
[12] R. Weegar and K. Sundström, “Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations,” PLoS One, vol. 15, p. e0237911, 2020.
[13] W. Lin et al., “Preoperative prediction of cervical cancer recurrence using explainable MRI radiomics,” Abdom. Radiol., 2025.

Biographies



Sneha Deshmukh is an Assistant Professor at Department of Computer Science and Engineering, MIT Art, Design and Technology University, Loni Kalbhor, Pune, India. with over 12 years of teaching experience. She is currently pursuing her Ph.D. and has research interests in Machine Learning, Artificial Intelligence, and Internet of Things (IoT).



Dr. Suvarna Pawae With 19+ years of teaching experience, she has served as a Head of Department for 15+ years, mentoring students and guiding UG/PG projects and research. She holds a Ph.D. in Artificial Intelligence and Machine Learning with expertise in IoT, prototype development, patents, copyrights, and industry-oriented solutions.



Dr. Sachin Arun Thanekar is an academician and researcher. His research interests include Artificial Intelligence, Machine Learning, Natural Language Processing, Explainable AI, and Air Quality Prediction. He has contributed to research publications, patents, academic projects, and industry-academia collaborations.



Prof. Komal Sunil Munde. She is currently working as an Assistant Professor at the Department of Computer Engineering, MITADT University. She has 9+ year experience. Her research areas include Machine learning, deep learning.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Automated Property and Risk Assessment System for Maharashtra

Pranav Divekar, Komal Shendage, Siddhi Kadam, Vedant Dongare & Prof. Prashant Sadaphule

Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India, pranavd4441@gmail.com, shendagekomal11@gmail.com, siddhikadam1803@gmail.com, vedantdongare10@gmail.com, prashant.sadaphule@aissmsioit.org

Abstract.

This paper presents an Automated Property and Risk Assessment System for Maharashtra that integrates government property records, OCR, computer vision, and machine learning for automated verification and risk assessment. Experiments achieved 96.8% extraction accuracy and 94.8% risk-classification accuracy, demonstrating improved efficiency, consistency, and traceability in property verification.

Keywords. Automated property verification, Maharashtra land records, IGR Maharashtra, OCR, YOLO, OpenCV, risk assessment, property intelligence, multilingual summarization.

1. INTRODUCTION

Property transactions in India require reliable verification of ownership, registration, and land records, but information is often fragmented across government portals and legal documents. This creates challenges for manual due diligence and increases the risk of inconsistencies and disputes [2, 5, 8, 11]. This study proposes a modular Automated Property and Risk Assessment System comprising data retrieval, document integration, multilingual summarization, and risk assessment. The system aims to convert heterogeneous property records into structured information and provide interpretable risk indicators, with the current implementation focusing mainly on automated data retrieval.

2. RELATED WORK

Recent research has applied deep learning, OCR, transformers, web extraction, and machine learning to document understanding, Legal AI, and anomaly detection [1, 3, 4, 6, 7, 9–11]. However, existing studies mainly address individual tasks. The proposed system integrates government-portal retrieval, OCR, document processing, multilingual summarization, and hybrid risk assessment into a unified property-verification workflow. It also supports 7/12 and mutation-record integration, OpenCV and YOLO-based extraction, and modular architecture for scalable property-risk analysis.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. PROPOSED SYSTEM

The system accepts inputs such as district, taluka, village, property identifier, and time range. These inputs pass through interconnected modules that progressively transform raw government data into structured records and risk indicators

3.1. System Architecture

The architecture includes four modules: Data Retrieval, Document Integration, Multilingual Summarization, and Risk Assessment. The retrieval module is implemented and tested, while the others are designed for future development.

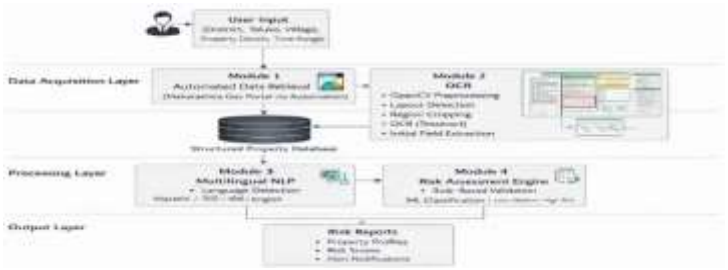


Figure 3.1.1: Workflow and data-flow architecture of the proposed automated property verification system

3.2. Data Retrieval

Selenium, OCR, and retries automate property-data retrieval and extract transaction details into a database.

3.3 Document Integration

7/12 records, mutation certificates, and transactions are combined using OCR, entity extraction, and fuzzy matching.

3.4 Multilingual Summarization

Transformer models generate summaries in English, Hindi, and Marathi.

3.5 Risk Assessment

Rule-based and machine-learning methods identify inconsistencies and classify properties as low, moderate, or high risk..

3.6 Technology Stack and Data Flow

The system uses Python, Selenium, Tesseract, OpenCV, YOLO, REST APIs, relational databases, and NLP models.

4. EXPERIMENTAL RESULTS

4.1 Experimental Setup

Experiments evaluated the retrieval module using 250 IGR Maharashtra property transactions across districts, document types, and registration years. The system used Python 3.11, Selenium, SQLite, REST APIs, and a CPU-based system with 16 GB RAM.

Performance was measured using extraction accuracy, completeness, OCR accuracy, risk-classification performance, and processing time.

4.2 Data Retrieval Results

Under dynamic web conditions, the retrieval module achieved the results in Table 2. The measurements indicate robust extraction and parsing performance across varying portal conditions.

Table 1: Performance metrics of the data retrieval module.

Metric	Value (%)
Extraction Accuracy	96.8
Record Completeness	95.4
Successful Query Rate	97.2
Data Parsing Accuracy	96.1

4.3 Region Detection and OCR

Computer vision segmented 7/12 records for OCR. Among 120 records, Tesseract provided fast and good multilingual recognition, while PaddleOCR had better detection accuracy.

Table 2: Comparison of OCR models.

OCR Model	Detection	Marathi	English	Speed
Tesseract OCR	Moderate	High	High	Fast
PaddleOCR	High	Moderate	High	Moderate
EasyOCR	Moderate	Low	Moderate	Slow

4.4 Risk Classification

The risk module was tested on 150 labeled property profiles. Table 4 shows strong performance, particularly for high-risk cases, attributed to rule-based validation and pattern recognition.

Table 3: Risk classification performance.

Risk Level	Precision (%)	Recall (%)	F1 (%)
Low	94.2	95.1	94.6
Moderate	91.3	89.8	90.5
High	96.8	97.4	97.1

Overall Accuracy: 94.8

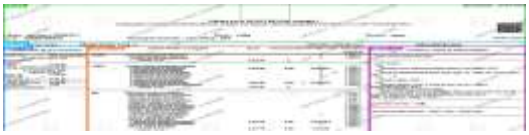


Figure 4.4.1: Region detection and segmentation results on a 7/12 document.

4.5 Efficiency and Scalability

Average processing time per property was 6.4 s for data retrieval, 4.1 s for document integration, 3.8 s for summarization, and 2.6 s for risk assessment, giving a total of 16.9 s. The results indicate a substantial reduction in processing time compared with manual verification and provide a traceable basis for scaling the workflow.

5. DISCUSSION

The results show that automated retrieval and OCR can effectively structure property records, although poor-quality and handwritten documents remain challenging. Combining rule-based and analytical methods improves high-risk property identification.

6. CONCLUSIONS

The proposed framework achieved 96.8% extraction accuracy and 94.8% risk-classification accuracy, demonstrating effective automated property verification. Future work will improve OCR, risk assessment, database scalability, and the user interface.

7. REFERENCES

[1] K. Boutalbi, F. Loukil, H. Verjus, D. Teliason and K. Salamanan, “Machine learning for text anomaly detection: A systematic review,” in Proc. IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1319–1324, 2023.

[2] Department of Land Resources, Government of India, “Digital India Land Records Modernization Programme (DILRMP),” 2023.

[3] Gupta, S. Furniturewala, V. Kumari and Y. Sharma, “Rhetorical role labeling of legal documents using transform-ers and graph neural networks,” arXiv:2305.04100, 2023.

[4] Majebraries, “Selenium web scraping: How to scrape dynamic sites step-by-step,” ScraperAPI Blog, 2025.

[5] Ministry of Law and Justice, Government of India, “National Judicial Data Grid Statistics,” 2024.

[6] P. Neelamegam and S. J. Nirmala, “Recent trends in Legal AI: A comprehensive review,” in Proc. ICAISS, pp. 152–159, 2025.

[7] S. K. Nigam, T. Dubey, G. Sharma, N. Shallum, K. Ghosh and A. Bhattacharya, “LegalSeg: Unlocking the structure of Indian legal judgments through rhetorical role classification,” in Findings of NAACL, pp. 1129–1144, 2025.

[8] NITI Aayog, “Land Records and Titles in India: Issues and Reforms,” Government of India, 2021.

[9] D. Premasiri, T. Ranasinghe, R. Mitkov, M. El-Haj and I. Frommholz, “Survey on legal information extraction: Current status and open challenges,” Knowledge and Information Systems, vol. 67, no. 12, pp. 11287–11358, 2025.

[10] D. Sporici, E. Cusnic and C.-A. Bolangiu, “Improving the accuracy of Tesseract 4.0 OCR engine using convolution-based preprocessing,” Symmetry, vol. 12, no. 5, p. 715, 2020.

[11] N. Subramani, A. Matton, M. Greaves and A. Lam, “A survey of deep learning approaches for OCR and document understanding,” arXiv:2011.13534, 2020.

Biographies



Pranav Divekar B.Tech. in Computer Engineering (2026), AISSMS Institute of Information Technology, Pune, with interests in AI, data processing, automation, and software development.



Vedant Dongare B.Tech. in Computer Engineering (2026), AISSMS Institute of Information Technology, Pune, with interests in AI, software development, data processing, and intelligent systems.



Siddhi Kadam B.Tech. in Computer Engineering (2026), AISSMS Institute of Information Technology, Pune, with interests in intelligent computing, data analysis, software engineering, and automation.



Komal Shendage received the B.Tech degree in Computer Engineering from the AISSMS Institute of Information Technology, Pune, India, in 2026. Her academic interests include computer engineering, intelligent systems, data processing, software development, and emerging technologies.



Prashant Shivling Sadaphule Assistant Professor of Computer Engineering at AISSMS Institute of Information Technology, Pune, with an M.E. in Computer Engineering and 17 years of academic and professional experience.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Explainable Deep Learning-Based Intrusion Detection System Using SHAP and LIME for Enhanced Transparency

Prasad A. Lahare[1], Babasaheb A. Abhale[2], Jagdish Kapadnis[3], Smita Thakare[4], Rupali Purkar[5], Rupali Rahinj[6]

[1,4]Asst. Prof. Information Technology, PVGCOE-SSDIOM, Nashik, prasadlahare7@gmail.com, skthakare29@gmail.com

[2]Asst. Prof. Computer Engineering, SND COE&RC, Yeola, atul.abhale@gmail.com

[3,5,6]Asst. Prof. Computer Engineering, PVGCOE-SSDIOM, Nashik, jagdishkapadnis@gmail.com, rupalipurkar172125@gmail.com, rupalirahinj007@gmail.com

Abstract.

Deep learning-based Intrusion Detection Systems (IDS) provide effective detection of complex cyber threats but often lack transparency and interpretability. This study proposes an interpretable IDS framework integrating deep learning with SHAP and LIME for global and local explanations. The framework is evaluated using NSL-KDD and CICIDS2017 datasets, achieving high detection performance while improving feature interpretability and model transparency. The proposed approach enhances decision validation, trust, and explainability, supporting applications in network security monitoring, Security Operations Centres, and anomaly detection.

Keywords. Explainable Artificial Intelligence (XAI), SHAP, LIME, Deep Learning, Intrusion Detection System (IDS), Cybersecurity, Anomaly Detection

1. INTRODUCTION

Modern networks face increasingly sophisticated cyber threats, making Intrusion Detection Systems (IDS) essential for network security. Deep learning improves attack detection but often lacks transparency. Explainable AI (XAI) techniques such as SHAP and LIME provide global and local explanations. This study proposes an interpretable IDS framework combining deep learning with SHAP and LIME, evaluated on NSL-KDD and CICIDS2017, to improve transparency without compromising detection performance and support practical cybersecurity applications.

2. BACKGROUND AND RELATED WORK

For intrusion detection many Deep learning models such as CNN, RNN, LSTM, and Bi-LSTM etc. have been analysed in a few research works. For example, based on benchmark datasets like NSL-KDD and CICIDS2017, investigators have shown that hybrid models that consist of RNN and Bi-LSTM have successfully captured both sequential attack and temporal dependencies patterns which improve detection accuracy. In a similar manner, by using CNN-based techniques for feature extraction have enhanced classification performance in high-dimensional network traffic data. Using the CICIDS2017 dataset, Chen

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

et al. (2024) introduced an intrusion detection method that integrates a Deep Neural Network (DNN) with the LIME (Local Interpretable Model-agnostic Explanations) technology [1]. This work's main contribution is the use of LIME to generate instance-level (local) explanations, which aid in understanding the model's classification of a particular network traffic instance as malicious or normal. As a result, DNN's individual predictions are easier to understand. But drawback of this method was lack of global interpretability. Using the UNSW-NB15 dataset, Singh et al. (2023) introduced an intrusion detection method that combines conventional Machine Learning (ML) models with SHAP (SHapley Additive explanations). This work's primary contribution is the use of SHAP for feature importance analysis, which allows for a thorough comprehension of how various input features affect the model's predictions [2]. By emphasizing the most important factors influencing intrusion detection conclusions, this increases transparency.

Table 1. Proposed System comparison with existing System

Author (Year)	**Method Used**	**Dataset**	**Key Innovation**	**Limitation**
Zhang et al. (2021)	CNN-based IDS	NSL-KDD	Improved feature extraction	No explainability
Kumar et al. (2022)	LSTM	CICIDS2017	High detection accuracy	Black-box model
Ali et al. (2023)	RNN + Bi-LSTM	NSL-KDD	Captures temporal patterns	Limited interpretation
Singh et al. (2023)	SHAP + ML	UNSW-NB15	Feature importance analysis	No deep learning
Chen et al. (2024)	LIME + DNN	CICIDS2017	Local explanation	No global insight
Proposed Work	DL + SHAP + LIME	NSL-KDD, CICIDS2017	Global + Local interpretability	—

3. KEY CONTRIBUTION

To enhance transparency in intrusion detection, an interpretable intrusion detection system that combines Explainable AI approaches (SHAP and LIME) with deep learning is proposed in this system. By delivering both fine-grained, instance-level explanations (using LIME for predictions) and global explanations (using SHAP for feature importance), multi-level model interpretability was attained. This ensured that there is no appreciable decline in classification efficacy, maintaining high detection performance and increasing model interpretability [5][6]. This method's robustness and improved generalization performance using benchmark datasets like NSL-KDD and CICIDS2017.

4. METHOD, EXPERIMENTS AND RESULTS

The suggested architecture offers a transparent intrusion detection system that integrates deep learning with interpretability methods to improve transparency and confidence. As shown in figure 1 at first, network traffic data from datasets like NSL-KDD and CICIDS2017 is introduced into the system, where it

undergoes preprocessing tasks such as data preprocessing, feature encoding and normalization to maintain data quality. The transformed data is subsequently sent through a feature engineering phase, where crucial features are chosen and modified for effective learning. The foundation of the system is a integrated deep learning model that consists of an RNN and Bi-LSTM, which effectively identifies temporal and sequential patterns in network traffic to identify it as normal or malicious. In the prediction of system, the explanation layer is implemented, which serves as the significant contribution of the model is its capability. SHAP delivers global interpretability by identifying the most significant features that affect the model throughout the full dataset, whereas LIME provides local interpretability by providing explanations for individual predictions through understandable surrogate models. Ultimately, the system provides not just the detection outcome but also provides valuable explanations, such as assessing the significance of features in classification outcomes and instance-specific insights, enhancing transparency, interpretability, and decision-making in cybersecurity contexts.



Figure 4.1: System Architecture for IDS using integrating Bi-LSTM with SHAP and LIME

5. RESULTS

The practical results show that proposed model achieves high efficiency and performance across all evaluation metrics. Proposed model attained an accuracy of 97.2%, outperforming LSTM (95.1%), CNN (93.5%), Random Forest (91.2%) and SVM (88.5%) shown in table 2.

Table 2 Performance Comparison of existing System proposed System

Model	Accuracy (%)	Precision	Recall	F1-Score	FPR
SVM	88.5	0.86	0.85	0.85	0.08
Random Forest	91.2	0.90	0.89	0.89	0.06
CNN	93.5	0.92	0.91	0.91	0.05
LSTM	95.1	0.94	0.94	0.94	0.04
Proposed Bi-LSTM + XAI	97.2	0.97	0.96	0.96	0.02

6. CONCLUSION

In this system we are using an explainable Intrusion Detection System (IDS) based on Bi-LSTM integrated with SHAP and LIME has been used to address the limitations of existing intrusion detection approaches. The proposed model leverages the strength of Bi-LSTM in

capturing bidirectional temporal dependencies in network traffic, resulting in improved detection of complex and previously unseen cyber-attacks. In this method we are evaluating standard benchmark datasets that demonstrate that in proposed system achieves high performance in terms of all feature like accuracy, precision, recall, and F1-score, while significantly reducing the false positive rate. These improvements highlight the effectiveness of the model in providing reliable and robust intrusion detection in proposed system.

7. REFERENCES

[1] X. Chen, Y. Liu, Z. Wang & H. Zhang (2024) "An efficient deep learning-based intrusion detection system using Bi-LSTM networks", IEEE Access, 12, pp. 12345–12358.

[2] Singh, R. Kumar & P. Sharma (2023) "A hybrid machine learning approach for intrusion detection using feature engineering", International Journal of Network Security, 25(3), pp. 345–356.

[3] M. Ali, S. Khan & T. Ahmad (2023) "A hybrid deep learning framework for intrusion detection using RNN-based architectures", IEEE Access, 11, pp. 56789–56802.

[4] C. Zhang, J. Li, N. Wang & D. Zhang (2025) "Research on intrusion detection method based on Transformer and CNN-BiLSTM in Internet of Things", Sensors, 25(9).

[5] P. Hermosilla, S. Berríos & H. Allende-Cid (2025) "Explainable AI for forensic analysis: A comparative study of SHAP and LIME in intrusion detection models", Applied Sciences, 15(13).

[6] U. Ahmed et al. (2024) "Enhancing intrusion detection performance using explainable ensemble deep learning", Scientific Reports.

Biographies



Prasad A. Lahare is an Assistant Professor in the Department of Information Technology at PVG's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India. His research interests include Deep Learning, Artificial Intelligence, and Cyber Security, with a particular focus on developing intelligent systems for threat detection and secure computing.



Babasaheb Abhale is an Assistant Professor in the Department of Information Technology at S. N. D. College of Engineering & Research Centre, Yeola, Nashik



Jagdish Kapadnis is an Assistant Professor in the Department of Computer Engineering at PVG's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India.



Smita Thakare is an Assistant Professor in the Department of Information Technology at PVG's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India.



Rupali Purkar is an Assistant Professor in the Department of Computer Engineering at PVG's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India.



Rupali Rahinj is an Assistant Professor in the Department of Computer Engineering at PVG's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

AI-Driven Tuberculosis Detection Using VGG-16

Prof. Dr. Reena Gunjan, Ms. Sharon Kasabe*, Mr. Devalh Anil Parmar,

Ms. Aadi Upadhyay, Mr. Hemant Sonewale

MIT Arts, Design and Technology University, Loni Kalbhor, Pune - 412201,

reena.gunjan@mituniversity.edu.in, kasabesharon@gmail.com, devalhparmar1@gmail.com, upadhyayaadi31@gmail.com, sonewaleh@gmail.com

Abstract.

Tuberculosis (TB) is still among the most essential infectious diseases that require immediate and precise detection to decrease mortality rates. Traditional methods are usually lengthy and require expertise. This research proposes an automatic method for TB detection based on the VGG-16 convolutional neural network utilizing chest X-rays (CXR). The proposed framework is implemented with TensorFlow and Keras and trained on open-source data. The experiments showed an accuracy rate of 77.8%. Nevertheless, relatively poor precision and recall reveal certain issues, such as class imbalance. The proposed application also incorporates a convenient, user-friendly interface with Gradio, allowing for instant predictions of X-rays uploaded by users.

Keywords. Tuberculosis, Deep Learning, CNN, VGG-16, Chest X-ray, TensorFlow, AI Diagnosis.

1. INTRODUCTION

TB remains a major worldwide health issue that severely impacts regions with limited medical resources, where diagnosis is delayed and inaccurate. Standard detection ways such as sputum microscopy and chest radiographs are tends to be slow, time- consuming, often has a lack of accuracy, are prone to human error, and need specialized expertise. Limited diagnostic resources in several regions' slow treatment in many areas, or delays due to workforce and the importance of AI driven approaches.

This approach can improve the efficiency of diagnosis and also lower the dependency on highly skilled medical experts. Enabling timely treatment and better patient outcomes, it provides faster results. Furthermore, the system promotes global healthcare equality by offering reliable diagnostic support in underdeveloped and resource-limited regions. Overall, the importance of integrating deep learning with medical imaging to achieve intelligent, accurate, and accessible Tuberculosis detection is highlighted by this approach.

2. RELATED WORK

Tuberculosis detection primarily works to utilize the CAD methodologies that are based on handcrafted edge and texture-based image features. Despite some of the automation showing

the dependence on manual features, these methods lacked flexibility and accuracy across the datasets. The advent of deep learning has advanced medical imaging by enabling the direct feature learning from raw data, and the models are trained in a way that they automatically learn complex features from X- ray. CNN architectures such as ResNet, DenseNet, and VGG are greatly enhanced for the classification accuracy for the thoracic conditions. In the case of Rajpurkar et al. (2017) [1].

CheXNet by Rajpurkar et al. was a key breakthrough or a mark as a major milestone, a 121-layer convolutional neural network (CNNs) was created for pneumonia detection that proved the abilities of deep learning architectures in clinical imaging. Moving on to this framework, Jaeger et al. [2] executed an early TB screening model integrating preprocessing and classification, serving as a reference for later studies. Showkatian et al. [3] improved TB detection through refined CNN integration and magnified data preprocessing, enhancing disease identification accuracy. Capellan-Martin et al. [4] developed a deep model that is small in size as well as very accurate yet is applicable in the real-time situation.

Prasad et al. [5] conducted comparative research between traditional machine learning and modern deep learning-based models with transfer learning that provides high-performance.

Overall, these articles indicate that the development of the traditional CAD systems into the new deep-learned-based-approaches is rather clear, and the potential of AI in TB detection is present. The incorporation of effective design models, such as GPU-enhanced systems, VGG models, and B user-friendly interface creates the basis of the usability of the system to create correct, accessible, and scalable TB diagnostic models to be utilized in clinical and distant health care processes. One paper presents a DH-DSU–optimized U-Net model for effective lesion segmentation in WCE images. It combines multiple feature extraction methods and uses an improved DNN for classifying inflammatory, polypoid, and ulcer diseases. Results show the proposed approach outperforms several existing optimization-based models.

3. METHODOLOGY

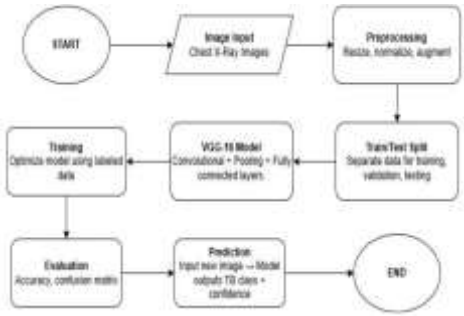


Fig. 3.1. Tuberculosis Detection Pipeline

TB is a serious bacterial disease that, for the most part, affects the lungs and demands timely detection for effective treatment. A deep learning model based on VGG16 is proposed for automated tuberculosis detection from chest X-ray images.The system is based on binary classification for TB-positive and TB-negative classes for chest X-ray images.Various preprocessing and augmentation methods are used for effective learning by the model. A web-based interactive system was also integrated for real-time prediction from uploaded images. The proposed framework demonstrates the possibility of AI-based TB detection in

real-world environments. A cost-effective diagnostic system was proposed for TB detection using Google Colab, TensorFlow, and Keras.

The general framework of the proposed approach for TB detection system is illustrated in **Figure 1**. As shown in the diagram, the process initiates with the input of CXR images, which are then subjected to necessary preprocessing steps such as normalization, augmentation, and resizing. The dataset is then classified into training, validation, and testing before passing through the VGG16 CNNs for feature identification along with categorization. The model is trained and optimized using labeled datasets, and the trained model is thereafter evaluated using accuracy metrics, ROC AUC, and a confusion matrix. Finally, the system provides instant diagnostic predictions, allowing new X-ray images to be analysed and classified as TB-positive or negative, along with the accuracy.

The presumed TB detection model automatically sorts CXR images into TB-positive or TB-negative classes. It is used in deep CNN particularly VGG16 architecture. The network has a layered design with fully connected stages, convolutional, and pooling. This structure helps in capturing low-level and high-level features of the lungs. As a result, it can detect the changes that are associated with TB infection

Before training, the image data went through several preprocessing steps that included resizing, normalization, random rotations, flips, and brightness and pixel adjustments to achieve better generalization and control overfitting. A user-oriented Gradio interface was integrated to simplify deployment, allowing direct upload of X-ray images for instant diagnostic predictions via the web. Although the framework supports transfer learning with pretrained CNNs, this implementation trains VGG16 from the ground up to extract features that are specifically relevant to tuberculosis pathology.

4. RESULT AND DISCUSSIONS

The model has achieved a total accuracy of 77.8, the preciseness = 14.9, recall = 7.1, and F1-score = 9.6 that detects Tuberculosis. It has accurately chosen the majority of the Normal images, i.e., (643/700), but with a poor sensitivity and a high specificity in terms of Tuberculosis cases (10/140). Therefore, this indicated lack of dataset balance and additional optimization to be made. Trained on the curated CXR dataset, the VGG16- based model demonstrated high competencies in distinguishing between the TB-positive and normal CXR. The reduction of the computational time was achieved without reducing the image resolution in the course of training. Random rotations and horizontal flips are data augmentation techniques that increase the ability of the model to generalize and retain high performance of unseen data depicted with assessed, unseen CXR images.

These results suggest that the VGG16 network successfully learns and identifies important radiographic patterns for TB detection, showing strong accuracy with both high sensitivity and specificity. Using the Gradio web interface, users are allowed to submit CXR images and direct prediction results of the trained model is more convenient and real-time clinical practices.

5. CONCLUSION

The study of this project concludes how the project actually works and reaches its goal. The correct use of VGG16 in order to understand and detect TB. It is seen how deep learning enhances TB detection and diagnosis using CXR images. The model learned well from the

images and could correctly tell the difference between TB and normal cases. Data augmentation methods, such as flipping as well as image rotation, has helped a lot in the case of model accuracy as well as its ability to work with new data . Some errors occurred with unclear or low-quality X-rays, which suggests that better image enhancement techniques could improve results further. A simple Gradio interface/app is also made where the doctors can upload an X-ray and instantly see if it shows “Tuberculosis-positive”. This system is mainly developed in order to help hospitals save time, reduce doctors’ workload, and make Tuberculosis screening easier even in places with fewer medical experts

6. REFERENCES

[1] P. Rajpurkar et al., “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays,”arXiv, 2017.
[2] S. Jaeger et al., “Automatic Tuberculosis Screening Using Chest Radiographs,” IEEE Trans. Med. Imaging, 2014.
[3] E. Showkatian et al., “Deep learning-based TB detection,” Polish Journal of Radiology,2022.
[4] D. Capellán-Martín et al., “Lightweight TB Detection Model,” IEEE ISBI, 2023.
[5] R. S. Prasad et al., “Comparative Study of TB Detection,” INDIACom, 2023

Biographies



Prof. Dr. Reena Gunjan received the Bachelor's and Master’s degree from JNV University, Jodhpur, Rajasthan and the philosophy of doctorate degree in Computer Engineering from Malviya National Institute of Technology, (MNIT) Jaipur, Rajasthan. She is currently working as a Professor in the Department of Computer Science and Engineering, School of Computing at MIT Art, Design, and Technology University, Pune. Her research areas include Medical Imaging, Artificial Intelligence and Machine Learning and Biomedical Signal Processing. She has been serving as a reviewer for many highly-respected journals.



Ms. Sharon Kasabe is a B.Tech. Computer Science Engineering student at MIT Arts, Design and Technology University, Pune. Her research focuses on Deep Learning, AI Driven Healthcare Diagnosis, and Computer Vision.



Mr. Devalh Anil Parmar is a B.Tech. Computer Science Engineering student at MIT Arts, Design and Technology University, Pune. His research focuses on AI diagnostics, Deep Learning, CNN architectures, and VGG-16 based models.



Ms. Aadi Upadhyay is a B.Tech. Computer Science Engineering student at MIT Arts, Design and Technology University. Her research interests focus on Artificial Intelligence, Machine Learning, Deep Learning, and intelligent healthcare applications.



Mr. Hemant Sonewale is a B.Tech. Computer Science Engineering student at MIT Arts, Design and Technology University. His interests focus on Artificial Intelligence, Deep Learning , AI-driven healthcare solutions, and intelligent diagnostic systems.

Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

SmartCry: Identifying Abnormal Infant Cry Patterns Using MFCC And Machine Learning Models

Dattatray Waghole[1], Smita Wagh[2] , Pranav Badole[3], Shrutam Hinge[4] ,Tanesh Devhare[5], Virendra Godse[6]

[1,2,3,4,5, 6]Department of Computer Engineering, JSPM's Jayawantrao Sawant College of Engineering Pune, India
dattatraywaghole@jspmjscoe.edu.in[1] , smitawagh@jspmjscoe.edu.in[2] , ppbadole28@gmail.com[3] , shrutamhinge9@gmail.com[4] , taneshdevhare07@gmail.com[5] , virendragodse04@gmail.com[6]

Abstract.

Analysis of infant cry patterns is an important approach for understanding various health conditions and needs of newborn babies. This research presents the system SmartCry for categorizing infant cry patterns into normal (hunger) and abnormal (disorders) categories, with main focus on a genetic disorder Cri Du Chat Syndrome. Mel Frequency Cepstral Coefficients (MFCC) is used for extracting various sound features and preprocessing techniques like normalization and segmentation is used to improve sound quality. Models such as Random Forest, CNN-LSTM are used to categorize different crying patterns. An embedded IOT device ESP32 is also used for real time detection. Thus, this system is designed to assist caregivers and parents an early detection of potential health issues.

Keywords. Infant Cry Patterns, MFCC, Machine Learning, Random Forest, CNN-LSTM, SmartCry, Cri Du Chat Syndrome.

1. INTRODUCTION

Infant crying communicates needs such as hunger, pain, and illness [1]–[3]. Abnormal crying patterns can also indicate genetic disorders such as Cri Du Chat Syndrome, characterized by a distinctive high-pitched, cat-like cry [4]. MFCC features and machine-learning models such as Random Forest are widely used for infant cry classification [5]–[7].

2. LITERATURE SURVEY

Research has widely applied acoustic features and machine learning to infant cry analysis [1], [2]. Random Forest has shown effectiveness in distinguishing normal and abnormal cries [3], while studies on Cri Du Chat Syndrome highlight abnormal crying as a potential early health indicator [4].

Table 1. Comparison of the system SmartCry with existing work on Infant Cry Analysis

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Algorithm	Cry Detection	Machine Learning used	Disease Detection
Infant Cry Review	Yes	Yes	No
Cry Classification	Yes	Yes	No
Genetic Disorder	No	No	Yes
MFCC analysis	Yes	Yes	No
This Work	Yes	Yes	Yes

3. PROPOSED SYSTEM / METHODOLOGY:

Audio Input Working: Crying audio is collected from the available datasets or from the generated synthetic datasets.

Pre-processing: The input files are processed by extracting all the noise and disturbances, using methods like normalization and segmentation.

Acoustic feature analysis: Various sound features such as pitch and frequency are captured from the collected input and are displayed in number format.

Model Training: Models such as Random Forest is trained to extract numerical values of the acoustic sound features to learn the pattern difference between the crying patterns.

Hardware Integration: After classification of crying patterns, an embedded system such as ESP32 is considered for practical implementation. Henceforth, ESP32 will relate to buzzer, and it will beep only when abnormal crying patterns are detected.

4. ARCHITECTURAL DIAGRAM

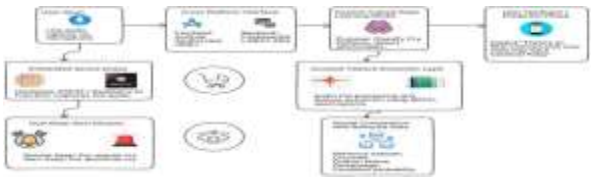


Figure 4.1. System Architecture of the proposed SmartCry system

The proposed architecture in Figure 1 focuses on automated infant cry classification into normal and abnormal crying patterns. The process starts from taking input either from a

microphone or dataset, where infant cry signals are captured which are processed further. Then we move on to the feature extraction stage, where MFCC features represent numerical values of the crying patterns. These features are provided to classification models like Random Forest. For practical implementation, an embedded system ESP32 is considered for real-time audio capture.

5. DATASET DESCRIPTION

Normal infant crying patterns like hunger, discomfort, and burping were obtained from standard datasets which were used in infant cry analysis research [1], [2]. Due to its rarity and limited availability of real-world abnormal cry data for rare genetic disorders such as Cri Du Cha, only a small number of abnormal cry samples are collected from medical and clinical data [3], [4]. To deal with this problem, we worked upon generation of synthetic data for abnormal cry patterns with the help of data augmentation techniques. The final dataset generated consists of both normal and abnormal crying patterns which will be further used for training and testing the classification models.

6. RESULT ANALYSIS

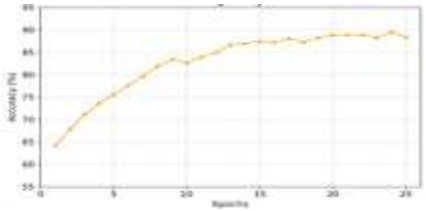


Figure 2: Accuracy vs Epochs Graph

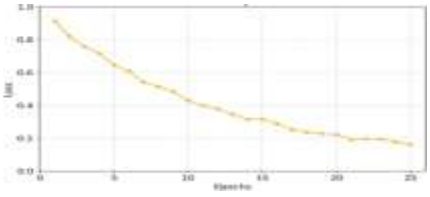


Figure 3: Loss vs Epochs Graph

The Fig 2 shows the accuracy of the model over different epochs. Here we can see that the accuracy keeps on increasing from 64% to 89% thus indicating that the model is effectively learning from the training dataset. The fig 3 shows the loss of the model during training. Here the loss is going from 0.9 to 0.16 hence indicating the prediction error is decreasing over the time.

7. CONCLUSIONS

The proposed system SmartCry shows an effective way of classification and categorization of infant cry patterns. The use of machine learning models increases the prediction capability by capturing both the sound as well as extracting the acoustic features by using Mel Frequency Cepstral Coefficients. The final outcome is a full working prototype, where the software ML model is integrated inside the embedded IOT device which is then placed inside a toy. The IOT device which is connected to a buzzer, will give an alert sound (beep

sound) only when the input audio data/file denotes an abnormal cry pattern (Cri Du Chat Syndrome). It won't give an alter sound for normal crying patterns (hunger, burping).

8. REFERENCES

[1] C. Ji, T. B. Mudiyanselage, Y. Gao, and Y. Pan, "A review of infant cry analysis and classification," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2021, no. 1, pp. 1–17, 2021, doi: 10.1186/s13636-021-00197-5.

[2] S. M. H. Hashemi, H. Kolivand, W. Khan, and T. Saba, "Infant cry analysis: A survey of datasets, features, and machine learning techniques," IEEE Transactions on Affective Computing, 2025, doi: 10.1109/TAFFC.2025.3636598.

[3] L. Liu, W. Li, X. Wu, and B. X. Zhou, "Infant cry language analysis and recognition: An experimental approach," IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 3, pp. 778–778, 2019, doi: 10.1109/JAS.2019.1911435.

[4] H. Bella, A. Ourrai, A. Radi, A. Hassani, A. Agadr, and R. Abilkassem, "Cri du Chat syndrome: A case study," SAS Journal of Medicine, vol. 10, no. 2, pp. 1–5, 2024, doi: 10.36347/sasjm.2024.v10i02.012.

[5] P. P. Singh and P. Rani, "An approach to extract feature using MFCC," IOSR Journal of Engineering, vol. 4, no. 8, pp. 21–25, 2014.

[6] R. Ranjan and A. Thakur, "Analysis of feature extraction techniques for speech recognition system," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 7C2, pp. 1–6, 2019.

[7] B. Vimal, M. Surya, D. Darshan, V. S. Sridhar, and A. Ashok, "MFCC based audio classification using machine learning," in Proc. Int. Conf. Computer Science and Engineering, India, 2020.

[8] D. Alagundi, S. Chikkamath, A. Patil, N. S. R., and V. Haridas, "Automatic infant cry classification using CNN architecture," in Proc. Int. Conf. Electronics and Communication Engineering, India, 2023.

[9] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," International Journal of Computer Science Issues, vol. 9, no. 5, pp. 272–278, 2012.

Biographies



Dr. Dattatray Waghole, Dean R&D at JSPMs Jayawantrao College of Engineering, PhD in Computer Engineering, with over a decade of experience in teaching with a focus on innovation through expertise in WSN, IoT and ML.



Prof. Smita Wagh, with over five plus years of experience in teaching at JSPMs Jayawantrao College of Engineering, with expiries in IoT, WSN and system programming, focused on innovation in new technologies.



Pranav Badole, 2026 pass out in Computer Engineering under SPPU University from JSPMs JSCOE, with expertise in ML, handling of IoT devices and application development.



Shrutam Hinge, bachelor's degree in computer engineering from JSPMs Jayawantrao College of Engineering, with expertise in Android and flutter development and spring boot working.



Tanesh Devhare, 2026 pass out in Computer Science from JSCOE, with expertise in integration of AI and ML models in wired/ wireless systems as well as skills in CAD integration with AI.



Virendra Godse, passed in Computer Engineering under SPPU University from JSPM College, with expertise in backend handling, java, firebase and deployments.

<Conference abbreviation>
<Conference Series name>
<Volume number and Year> <DOI Number>

AI Based Smart And Sustainable Biogas System For Rural Empowerment

Dattatray Waghole[1] , Smita Wagh[2] , Shreetej Korde[3] , Joel Chandanshiv[4] , Shreyas Aher[5]

[1.2.3.4.5]Department of Computer Engineering
JSPM's Jayawantrao Sawant College of Engineering Pune, India
[1]dattatraywaghole@jspmjscoe.edu.in, [2]smitawagh@jspmjscoe.edu.in,
[3]shreetejkorde@gmail.com, [4] joelchandanshiv@gmail.com, [5]shreyasaher1534@gmail.com

Abstract.

The AI-Based Smart Biogas System automatically converts organic waste into renewable biogas and nutrient-rich biofertilizer. Sensors and microcontrollers monitor temperature, pH, and nutrient levels, while AI algorithms optimize anaerobic digestion. Compared with traditional systems, it improves biogas yield by approximately 30%. The biogas is purified through scrubbing for cleaner energy use, while the treated and oxygenated slurry serves as an enhanced biofertilizer. An IoT dashboard enables real-time monitoring and control, promoting clean energy, sustainable agriculture, and improved efficiency in rural areas.

Keywords: Smart Biogas System, AI-Based Automation, Anaerobic Digestion, IoT Monitoring, Renewable Energy, Biofertilizer, Rural Sustainability, Waste-to-Energy.

1. INTRODUCTION

Anaerobic digestion sustainably converts organic waste into methane-rich renewable fuel and nutrient-rich biofertilizer. It is especially valuable in rural areas with abundant organic waste but limited energy access, addressing energy shortages, waste management, and environmental concerns [1], [2]. The resulting slurry also supports sustainable agriculture by reducing reliance on artificial fertilizers [3].Traditional biogas systems depend on manual monitoring and lack effective control of temperature, pH, gas composition, retention time, and feedstock variations, resulting in unstable microbial activity and inconsistent methane yield [4], [5]. IoT and AI enable real-time monitoring, prediction, and automated control [6]. Sensors collect process data, while LSTM and XGBoost models predict gas yield, detect abnormalities, and optimize digestion [7], [8].The proposed system integrates AI, IoT, water treatment, UV disinfection, and nutrient enrichment. It aims to increase methane production by approximately 30%, improve fertilizer quality, and provide remote monitoring through an Android dashboard [9]. This self-regulating solution promotes decentralized renewable energy, sustainable agriculture, and rural empowerment.

2. LITERATURE SURVEY

IoT-based biogas monitoring systems improve efficiency by enabling real-time tracking of temperature, pH, and gas production [1], [2]. However, they mainly focus on data collection and often lack predictive capabilities. Machine-learning techniques address this limitation by modelling complex relationships between input parameters and biogas yield, thereby improving system accuracy and performance [9]. Time-series models such as LSTM

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

effectively predict biogas production trends, while ensemble-learning methods such as XGBoost provide reliable classification and regression results [3], [10]. AI-based optimization further improves anaerobic digestion by enabling better control of process parameters [8]. IoT-enabled systems have also been integrated with agricultural applications to promote the use of biogas slurry as an organic fertilizer [4].

2. PROPOSED SYSTEM/METHODOLOGY

- **Input and Mixing:** Organic waste is collected and mixed to form a uniform slurry.
- **Monitoring and Digestion:** Sensors monitor the digester's temperature, pH, and turbidity to support efficient biogas production.
- **Biogas Purification:** The produced biogas is water-scrubbed to obtain cleaner methane.
- **UV Disinfection:** UV light removes harmful bacteria from the residual slurry.
- **Enrichment:** Oxygenation and continuous NPK monitoring improve the slurry's nutrient content.
- **Output Processing:** The enriched slurry is dried or pelletized into nutrient-rich biofertilizer.
- **Monitoring and Control:** An IoT-based Android application enables real-time monitoring and control.

3. ARCHITECHTURE

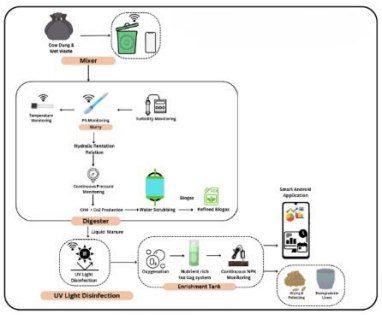


Figure 1. System Architecture

The proposed AI-based smart biogas system converts cow dung, kitchen waste, and agricultural residues into methane-rich biogas and biofertilizer. An anaerobic digester uses sensors to monitor temperature, pH, turbidity, and pressure, while LSTM and XGBoost predict gas yield and control system components. Water scrubbing improves methane quality. Digestate undergoes UV disinfection, oxygenation, and nutrient monitoring for fertilizer production. An IoT-enabled Android dashboard supports remote monitoring and control, providing a scalable solution integrating waste management, renewable energy, AI, and sustainable agriculture.

4. MATHEMATICAL MODEL

The following is a mathematical representation of the system:

Sensor Data Acquisition:

$$S(t) = [T(t), pH(t), G(t), N(t)] \quad (1)$$

Preprocessing:

$$S' = \psi(S) \quad (2)$$

Prediction Model (Hybrid):

$$Y = f\ LSTM(S') + f\ XGB(S') \quad (3)$$

Output Optimization:

$$\boldsymbol{O = g(Y)} \quad (4)$$

Where: S(t) = Sensor data (temperature, pH, gas, nutrients), S' = processed data , Y = predicted biogas yield/efficiency ,O= optimized output (biogas + biofertilizer).

5. RESULT ANALYSIS

The proposed system addresses conventional biogas limitations through IoT- and AI-based monitoring and control. Real-time sensors monitor key parameters, while LSTM and XGBoost models predict biogas yield and improve process efficiency. Purification and UV disinfection enhance biogas and biofertilizer quality, while the IoT dashboard enables remote monitoring and automation. However, system performance depends on sensor accuracy and data quality.

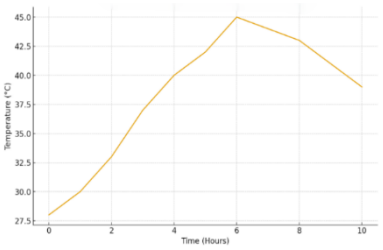


Figure 3. Temperature vs Time Graph

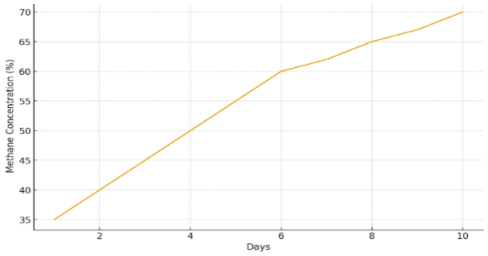


Figure 4. Methane Concentration vs Days Graph

		Sum of Squares	df	Mean Square	F	Sig.
VAR00001	Between Groups	533173.867	2	266586.933	15878.619	<.001
	Within Groups	83911.672	4998	16.789		
	Total	617085.538	5000			
VAR00002	Between Groups	123614.137	2	61807.069	650.585	<.001
	Within Groups	474821.625	4998	95.002		
	Total	598435.762	5000			
VAR00003	Between Groups	8017.269	2	4008.634	183.532	<.001
	Within Groups	109164.256	4998	21.842		
	Total	117181.524	5000			
VAR00004	Between Groups	8.593	2	4.296	18.710	<.001
	Within Groups	1147.698	4998	.230		
	Total	1156.291	5000			
VAR00005	Between Groups	983.143	2	491.572	19.088	<.001
	Within Groups	128716.438	4998	25.754		
	Total	129699.581	5000			

Table 2. Bayesian Correlation Analysis for a Smart and Sustainable Biogas System Using AI

6. CONCLUSION

The conclusion from the above system description states that it is an AI-based smart biogas system that combines IoT, LSTM, and XGBoost for real-time monitoring and optimization,

presented in this paper. The technology further ensures improved process stability and output quality while increasing biogas production efficiency by about 30%. Purification, disinfection, and enrichment procedures are added to improve the overall sustainability and efficacy of the system. The future work focuses on expanding the dataset with diverse real-world scenarios, integrating multi-sensor data, developing lightweight edge-AI models, and implementing adaptive learning for continuous system improvement.

7. REFERENCES

[1] G. Y. N. V. and R. R., "Hardware implementation of anaerobic biogas digester for household and community purposes," in 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), Kuala Lumpur, Malaysia, 2023, pp. 1–8. doi: 10.1109/i-PACT58649.2023.10434565.

[2] T. Tenzin et al., "Optimization of biogas production from kitchen waste installed at College of Science and Technology," in 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2018, pp. 956–960. doi: 10.1109/GUCON.2018.8674963.

[3] N. U. Okafor and D. T. Delaney, "Application of machine learning techniques for the calibration of low-cost IoT sensors in environmental monitoring networks," in 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), New Orleans, LA, USA, 2020, pp. 1–3. doi: 10.1109/WF-IoT48130.2020.9221246.

[4] Hajji, N. Laaroussi, M. Rhachi, and M. Garoum, "The effect of light on anaerobic digestion process of household and similar waste under mesophilic regime," in 2021 9th International Renewable and Sustainable Energy Conference (IRSEC), Morocco, 2021, pp. 1–3. doi: 10.1109/IRSEC53969.2021.9741135.

[5] R. Mouleeshuwarapprabu et al., "Design and development of biogas purifier," in 2023 International Conference on Energy, Materials and Communication Engineering (ICEMCE), Madurai, India, 2023, pp. 1–5. doi: 10.1109/ICEMCE57940.2024.10433973.

[6] S. Khan, "Nanotechnology applications in food and agriculture," in 2017 IEEE 17th International Conference on Nanotechnology (IEEE-NANO), Pittsburgh, PA, USA, 2017, pp. 877–878. doi: 10.1109/NANO.2017.8117471.

[7] Pal and G. S. Ilango, "Biogas resource estimation for different temperature zones in India using a thermal model," in 2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2025, pp. 1–5. doi: 10.1109/SCEECS64059.2025.10941615.

[8] S. Swami, S. Suthar, R. Singh, A. K. Thakur, L. R. Gupta, and V. S. Sikarwar, "Integration of anaerobic digestion with artificial intelligence to optimise biogas plant operation," Environment, Development and Sustainability, vol. 27, no. 5, pp. 9773–9803, May 2025. doi: 10.1007/s10668-023-04326-2.

[9] J. Zou et al., "Machine learning for enhancing prediction of biogas production in full-scale dry anaerobic digestion," Renewable Energy, 2024. doi: 10.1016/j.renene.2024.03.1761.

[10] S. Peng et al., "Biogas production prediction based on feature selection and ensemble learning," Applied Sciences, vol. 14, no. 2, p. 901, 2024. doi: 10.3390/app14020901.

Biographies



Dr. Dattatray Waghole is the Dean of Research and Development at JSPM's JSCOE, Pune. He has 13 years of experience, 70+ publications, and 16+ patents. His interests include AI, ML, IoT, WSN, and quantum computing.



Prof. Smita Wagh is an Assistant Professor at JSPM's JSCOE, Pune, and is pursuing a Ph.D. in CSE. She has 6+ years of academic experience, 15+ publications, and two international patents. Her interests include AI, machine learning, and blockchain technology.



Shreetej Korde holds a bachelor's degree in Computer Engineering from JSPM's JSCOE, Pune. His interests include AI, machine learning, IoT, and intelligent systems. Winner of National & International Hackathon with experience in innovative research projects.



Joel Chandanshiv holds a bachelor's degree in Computer Engineering from JSPM's JSCOE, Pune. His interests include cloud computing, DevOps, AI and IoT. Winner of National & International Hackathon with experience in innovative research projects.



Shreyas Aher holds a bachelor's degree in Computer Engineering from JSPM's JSCOE, Pune. His interests include Flutter development, AI, Cloud Computing. Winner of National & International Hackathon with experience in innovative research projects.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

AirGuard: Micro-Doppler-Based Target Classification Using Edge AI

Dattatray Waghole[1],Smita Wagh[2],Arya Dongare[3],Anjum Maner[4],Kirti Chandratike[5]

Department of Computer Engineering, JSPM's Jayawantrao Sawant College of Engineering Pune, India
dattatraywaghole@jspmjscoe.edu.in[1] , smitawagh@jspmjscoe.edu.in[2] , aryadongare20@gmail.com[3] , anjum1662004@gmail.com[4] , kirtichandratike@gmail.com[5]

Abstract.

In recent years, identifying the difference between drones and birds has become increasingly difficult in low-altitude airspace monitoring systems. Therefore, the study proposed as a micro-Doppler-based approach that was combined with deep learning techniques in order to address these challenges. Here, the radar signals have been transformed into time-frequency representations by using the Short-Time Fourier Transform (STFT), which is used to capture unique motion characteristics like rotor movement and wing flapping. These spectrograms are then used to train a YOLOv11-based model to support a real-time classification process. The proposed system was observed to achieve approximately 96% accuracy while also maintaining low latency parameters, making it suitable for deployment purposes on edge devices. Overall, these approaches are seen to provide an efficient and scalable solution supporting real-time aerial object classification.

Keywords. Micro-Doppler Radar, Edge AI, YOLOv11, Drone Detection, Bird Classification, STFT, Spectrogram Analysis, Real-Time Target Classification.

1. INTRODUCTION

The rapid growth in the usage of drones has thereby made an airspace monitoring system to seem more complex. One of the key challenges among all of them is the differentiation of drones from birds, as both of them are seen to generate similar radar signatures.[1][8]. Traditionally the techniques like Radar Cross Section (RCS) were used, which often failed to provide reliable results hence leading to misclassification [3][7]. In order to overcome this limitation as a solution micro-doppler signatures have been utilized to capture fine-grained motion patterns also including rotor dynamics in drones and wing movements in birds effectively [1]. The recent advancements in deep learning have further improved classification accuracy by the convergence of radar signals into spectrogram images therefore allowing the automatic feature learning process[5][9][10].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE SURVEY

Many researchers were found to look into the micro-Doppler techniques for the purpose of classifying aerial targets. Where the earlier studies demonstrated that these signatures could effectively distinguish drones from birds using the radar data [1]. Along with deep learning performance has improved significantly, especially when the radar signals have been converted into spectrogram images also making the feature extraction step easier [9]. Models like Convolutional Neural Networks (CNNs) have been popular because they are used to extract important patterns from these spectrograms [8]. Recently, YOLO-based models have gained attraction among all due to their capabilities like speed and real-time detection [10]. The Data augmentation techniques has also helped to improve model accuracy and thereby reduce overfitting [7]. Despite of all these advancements the challenges remained such as performance drop in the noisy conditions and the limitations with small datasets [6][5]. Thus, a more robust and efficient system was still needed for reliable real-time aerial target classification.

3. PROPOSED SYSTEM / METHODOLOGY

Radar Signal Working: Radar signals of the and birds are collected from the available datasets or radar-based sensing systems.

Pre-processing: The collected radar signals are processed in order to reduce the noise and disturbances using the methods such as normalization and segmentation.

Spectrogram Analysis: The pre-processed radar signals are then converted into spectrograms using the Short-Time Fourier Transform (STFT) to capture the micro-doppler patterns such as rotor movement in drones and the wing movements in birds.

Model Training: YOLOv11 is trained using the generated spectrogram images to automatically learn the distinguishing patterns between drone and bird signatures and then perform target classification.

Edge AI Integration: After the classification of the aerial targets is done, the trained model is then considered for the deployment on an edge or mobile device. Then the system will provide fast and low-latency real-time classification of drones and birds.

4. ARCHITECTURAL DIAGRAM

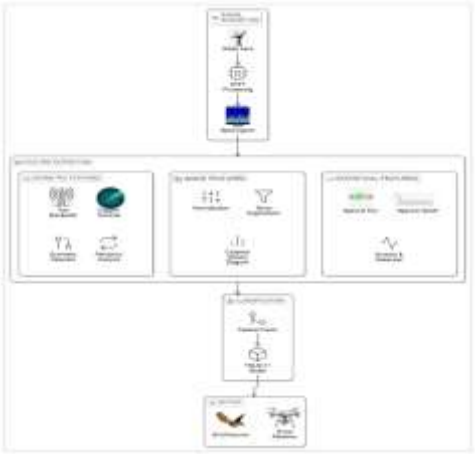


Figure 4.1. System Architecture of the proposed micro-doppler system

The above system first processes to collects raw radar signals and then clean them by removing noise in order to enhance quality. These signals are then converted into spectrograms using the STFT, which helps in extraction of useful features. These features are then applied as input to the YOLOv11 model for real-time classification.

5. DATASET DESCRIPTION:

Radar signatures of drones and birds were obtained from standard radar datasets used in aerial target classification and micro-Doppler analysis research [1], [3]. The collected radar data represents different motion patterns of drones and birds, including rotor movements and wing-flapping patterns.Due to the limited availability and variability of real-world radar samples, data augmentation techniques were used to increase the diversity of the available data. Techniques such as time shifting, noise addition, and scaling were applied to generate additional samples [7].

6. MATHEMATICAL MODEL:

Mathematical representation of our system:

The STFT is represented as:

$$X\,(t,f)\;=\;\int\; x(\tau)\,w\,(\tau\;-\;t)\,e\text{^}(-j2\pi f\tau)\,d\tau \tag{1}$$

Feature extraction function:

$$F\;=\;\phi(X) \tag{2}$$

The output class:

$$Y\;=\;f\,(F;\,\theta) \tag{3}$$

X(t) = Input raw radar signal, X′ = Pre-processed radar signal , X(t,f) = STFT-generated spectrogram, F = Extracted micro-Doppler features from the spectrogram, Y = Model prediction (Drone or Bird), θ = Trained YOLOv11 model parameters.

7. RESULT ANALYSIS:

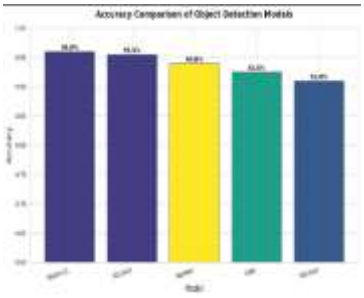


Figure 2: Accuracy Comparison



Figure 3: Application Interface

8. CONCLUSIONS:

In this work we have presented the micro-Doppler-based system for classifying drones and birds utilizing STFT and a YOLOv11 model. The results are seen to promise us with the model achieving approximately 96% accuracy while supporting the real-time edge deployment. Using the spectrogram-based features we improved overall classification performance. Nonetheless, there is still some room for the enhancement, especially with the noisy data and the unfamiliar situations.

9. REFERENCES:

[1] Rahman, Samiur, and Duncan A. Robertson. “Radar Micro-Doppler Signatures of Drones and Birds at K-Band and W-Band” appeared in Scientific Reports on November 26, 2018.

[2] “YOLOv11: An Improved Real-Time Object Detection Model” was released as an arXiv preprint in 2024.

[3] “Classification and Recognition of Objects on Radar Portraits Formed by the Equipment of Mobile Small-Size Radar Systems.”ieee.org, September 29, 2024.

[4] “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” Alec Radford and co-authors describe a GAN-based approach for unsupervised feature learning. Chen, J., X. Ran. 2019. “Deep Learning with Edge Computing: A Review.” Proceedings of the IEEE 107 (8): 1655–1674.

[5] Jürgen Schmidhuber’s paper “Deep Learning in Neural Networks: An Overview” appeared in Neural Networks, Volume 61, in 2015.

[6] Shorten, Connor, and Taghi M. Khoshgoftaar. 2019. “A Survey on Image Data Augmentation for Deep Learning.” The study appears in the Journal of Big Data, Volume 6, Issue 1, published as Article 60.
[7] “Radar System Engineering: From Requirements to Deployment.” mathworks.com, September 29, 2024
[8] Ulzhalgas Seidaliyeva. “Radar-Based Drone Detection Technologies.” On January 4, 2024.
[9] D. Bhatt et al., “CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope,” Electronics, vol. 10, no. 20, p. 2470, 2021. Available:
[10] N. Kale, S. N. Gunjal, M. Bhalerao, H. E. Khodke, S. Gore, and B. J. Dange, “Crop yield estimation using deep learning and satellite imagery,” International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 10s, pp. 464–471, 2023.

Biographies



Dr. Dattatray Waghole, Dean R&D at JSPMs Jayawantrao College of Engineering, PhD in Computer Engineering, with over a decade of experience in teaching.



Prof. Smita Wagh, with over five plus years of experience in teaching at JSPMs Jayawantrao College of Engineering.



Arya Dongare, 2026 pass out in Computer Engineering under SPPU University from JSPMs JSCOE.



Anjum Maner, 2026 pass out in Computer Engineering under SPPU University from JSPMs JSCOE.



Kirti Chandratike, 2026 pass out in Computer Science from JSCOE, with expertise in API development and MongoDB.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Speak My Scan: An Assistive Image-to-Speech System for Visually Impaired Users

Supriya Jagtap[1], R. Bhutada[2], P. Bangar[3], R. Bhusare[4], K. Akolkar[5], R. Biradar[6]

[1,2,3,4,5,6]Assistant Professor, Information Technology Department, Vishwakarma Institute of Technology, Pune, India Information Technology Department, Vishwakarma Institute of Technology, Pune, India

[1]supriya.jagtap@vit.edu , [2]riddhi.bhutada24@vit.edu, [3]pranali.bangar24@vit.edu, [4]rohan.bhusare24@vit.edu, [5]kartik.akolkar24@vit.edu, [6]ritesh.biradar24@vit.edu

Abstract.

Visually impaired users have not yet access to printed information, as the conventional screen readers are not able to directly interpret the content of images or of a scanned document. To solve this issue, this paper introduces Speak My Scan, an offline assistive system based on Optical Character Recognition (OCR) and Text-to-Speech (TTS) technologies which converts image and document text to speech. The system proposed is based on OpenCV for image preprocessing, Tesseract for the extraction of information in the form of text, and pyttsx3 for offline speech synthesis. It accepts image files, multi page PDF and live camera input via a GUI. A test set of 50 images of printed documents, book pages, notices and signage were used in the experimental evaluation under various lighting and image conditions. For printed documents in good lighting, the system achieved a top scorer of over 94% OCR accuracy and adaptive preprocessing improved OCR accuracy by about 11 percentage points over the direct processing of raw images. The speed of the processing time of a single A4 page was 2.3 seconds on an average laptop. The suggested system shows that an offline OCR-TTS system can be practical and accessible for providing reading assistance, even without special equipment or Internet access.

Keywords. Optical Character Recognition, Text-to-Speech, Assistive Technology, Image Processing, Computer Vision, Visually Impaired

1. INTRODUCTION

Printed information is a difficult resource to access for the visually impaired. Traditional screen readers work well with digital text on computer screens, but can't access information in pictures, books, notices or scanned documents. Consequently, many of the print materials that are readily available are inaccessible to visually impaired people. OCR can be used to convert an image to text, and once the text is extracted from an image, it can be converted into speech using TTS. These technologies can be used together in an OCR-to-speech pipeline to enable access to printed information by speech output. The Speak My Scan system consists of adaptive image preprocessing, Tesseract OCR and pyttsx3 speech synthesis by a single desktop application. The proposed solution is not cloud dependent, nor requires any special hardware to be installed. It also can accept various input types such as

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

static images, PDF files and live camera input. The goal of the system is to give visually impaired users an easy, low-cost and convenient alternative access to the printed information..

2. RELATED WORK

There are a number of researches on OCR and TTS for assistive applications. Smith developed one of the popular open-source OCR systems, Tesseract OCR, and OpenCV offers image-processing methods to enhance the image quality prior to OCR. Ben Atitallah built an on-the-fly TTS system which required special hardware, but it worked with Raspberry Pi. Ariyoshi and Chun suggested a cloud-based processing scheme based on machine learning for OCR-to-audio recognition, which has a high recognition accuracy, but is not applicable in areas with weak Internet connectivity. Al-Mansoori studied on OCR and generation of braille using IoT, and Raj et al. implemented OCR based reading system using LabVIEW. Other research works explored the design of the multilingual image-to-speech systems and the newspaper reading application with the aid of OCR and layout analysis. The Speak My Scan proposed is a lightweight open source desktop application that integrates preprocessing, OCR, pdf processing, camera input and offline TTS.

3. PROPOSED INTEGRATED SYSTEM AND METHODOLOGY

The Speak My Scan proposed system is composed of three modules, namely Input Acquisition, Image Preprocessing, OCR-Based Text Extraction, Text Post-Processing and Speech Synthesis, which take the visual input and convert it into audible speech. The stages are separately upgradeable as well as replaceable.

3.1 Input Acquisition

Here are three types of input supported by the system: static images (JPEG, PNG, BMP), multi-page PDF documents converted to images by pdf2image and real-time camera input (OpenCV). This makes a system that can read documents, books, notices and signs.

3.2 Image Preprocessing

The image is enhanced with OpenCV prior to OCR. The pipeline consists of several operations: grayscale conversion, 5×5 Gaussian blur, adaptive Gaussian thresholding with a block size of 11, and morphological dilation with a kernel of size 2×2. These operations dampen noise, deal with uneven lighting and increase the visibility of characters. 50 document images were used to fine-tune the parameters.

3.3 OCR-Based Text Extraction

Tesseract OCR v5.x is fed with the preprocessed image via pytesseract. The system adopts the LSTM based OCR engine with PSM 6, which is appropriate for document-like text. English is default language, other languages can be added via language data files.

3.4 Text Post-Processing

Non printable characters, normalizes whitespace and removes isolated character artefacts using regular-expression rules are cleaned when creating raw OCR output. The text is next cleaned and then presented in the GUI for speaking.

3.5 Speech Synthesis

The cleaned text is then converted to speech with pyttsx3 using the OS's built-in speech synthesizer and without requiring any internet connection. Speech rate, voice and volume can be set, and TTS works in a separate thread to maintain responsiveness of the GUI.

4. RESULTS AND DISCUSSION

50 images of various types of printed documents, book pages, notices and signage were used in various conditions to test Speak My Scan, including good lighting, dim lighting and slight skew.

4.1 OCR Performance

It obtained a higher OCR accuracy of over 94% at the character level for the printed documents under good lighting conditions. Adaptive thresholding was used to enhance the accuracy of the OCR by about 11 percentage points over the raw images processed straight to Tesseract.

4.2 PDF and Speech Processing

The preprocessing pipeline was able to extract high-fidelity text from the PDFs generated from the digital files and the scanned PDFs. Understandable speech was produced by the pyttsx3 engine, and preliminary testing with five visually impaired volunteers resulted in acceptable speech understanding. A laptop computer using an Intel Core i5 processor with 8 GB RAM was used to measure the processing time for a single page of A4: 2.3 seconds, adequate for assistive use.

4.3 Limitations

Low-quality images -- especially those below about 150 DPI -- and handwritten or ornamental fonts suffer in the quality of the recognition performance. Documents with multiple columns can also cause reading order issues if no further layout analysis is performed.

The default mode of the current system is English; other languages will need to use the appropriate language packages of the Tesseract software.

5. CONCLUSION

Speak My Scan is an offline and open-source assistive system presented in this paper to translate text from images and documents to audible speech for visually impaired users. The system features pre-processing using OpenCV, OCR using Tesseract, and then post-processing the text and using pyttsx3 for the TTS through a Tkinter GUI. Works with images, PDFs, and camera input—no Internet connection or special equipment. The system gained OCR accuracy of over 94% from printed documents under good illumination, and adaptive preprocessing increased the accuracy of OCR by about 11 percentage points. The average processing time per A4 page was 2.3 seconds and preliminary testing indicated good speech

comprehension. Plans involve handwritten and low-quality text recognition, multilingual support, deep learning-based OCR, document layout analysis, processing of video streams, deployment on low-cost hardware like Raspberry Pi, and more.

6. REFERENCES

[1] Ben Atitallah, "Development of Real-Time TTS System Using Raspberry Pi for Visually Impaired," in Proc. Int. Conf. on Innovative Trends in Electronics and Communication Engineering (ICITECE), Bangalore, India, 2024, pp. 1–6.

[2] J. Ariyoshi and J. Chun, "Machine Learning-Based OCR for Text-to-Audio Conversion," IEEE Access, vol. 13, pp. 1–12, 2025, doi: 10.1109/ACCESS.2025.0000001.

[3] D. Raj, H. F. Tazeen, V. Ravish, A. H. N. Anupama, and R. N. Nagashree, "An Assistive Reading System for Visually Impaired and Blinds Using OCR and TTS Techniques on LabVIEW," Int. J. Eng. Res. Technol. (IJERT), vol. 7, no. 10, pp. 1–5, Jun. 2019.Al-Mansoori, "IoT-Powered OCR-Braille System for Blind and Deaf-Bl

[4] ind Users," IEEE Internet Things J., vol. 11, no. 8, pp. 1–10, 2024.

[5] R. Smith, "An Overview of the Tesseract OCR Engine," in Proc. 9th Int. Conf. Document Analysis and Recognition (ICDAR), Parana, Brazil, 2007, pp. 629–633, doi: 10.1109/ICDAR.2007.4376991.

[6] Sundar, O. Abinaya, and V. Ajithkumar, "An Optical Character Recognition Framework Based Newspaper Reader Application for Blind," in Proc. 2021 Int. Conf. Computing, Communication, and Intelligent Systems (ICCAP), Greater Noida, India, Dec. 2021, pp. 1–6, doi: 10.4108/eai.7-12-2021.2314975.

[7] R. Rosalina, H. Fahmi, and G. Sahuri, "Artificial Intelligence Multilingual Image-to-Speech for Accessibility and Text Recognition," Int. J. Artif. Intell., vol. 14, no. 1, pp. 1–10, 2025.

[8] S. Chaturvedi, M. Chechani, K. Gautami, and K. Joshi, "System to Provide Reading Aid to Visually Impaired People," Asian J. Comput. Technol., vol. 9, no. 2, pp. 1–8, 2021.

[9] Sharma, A. Srivastava, and A. Vashishth, "An Assistive Reading System for Visually Impaired Using OCR and TTS," Int. J. Comput. Appl., vol. 95, no. 2, pp. 13–18, Jun. 2014, doi: 10.5120/16566-6231.

[10] P. Langde, S. Patil, and P. Langde, "Automating Document Narration: A Deep Learning-Based Speech Captioning System for Visually Impaired Person," Int. J. Intell. Syst. Appl. Eng. (IJISAE), vol. 12, no. 2, pp. 1130–1140, 2024, doi: 10.18201/ijisae.2024.7469.

[11] G. Bradski, "The OpenCV Library," Dr. Dobb's J. Softw. Tools, vol. 25, no. 11, pp. 120–125, Nov. 2000.

Biographies



Prof. Supriya Jagtap is currently working as an Assistant Professor in the Information Technology Department at Vishwakarma Institute of Technology, Pune, India. Her professional interests include teaching, student mentoring, and academic activities in the field of Information Technology.



Riddhi Bhutada is a student in the Information Technology Department at Vishwakarma Institute of Technology, Pune, India. Her interests include Artificial Intelligence, Machine Learning, Data Science, and software development.



Pranali Bangar is affiliated with Vishwakarma Institute of Technology, Pune, India, as a student in the Information Technology Department. Her interests include Artificial Intelligence, Machine Learning, Data Science, and software development.



Rohan Bhusare is affiliated with Vishwakarma Institute of Technology, Pune, India, as a student in the Information Technology Department. His interests include Artificial Intelligence, Machine Learning, Data Science, and software development.



Kartik Akolkar is affiliated with Vishwakarma Institute of Technology, Pune, India, as a student in the Information Technology Department. His interests include Artificial Intelligence, Machine Learning, Data Science, and software development.



Ritesh Biradar is affiliated with Vishwakarma Institute of Technology, Pune, India, as a student in the Information Technology Department. His interests include Artificial Intelligence, Machine Learning, Data Science, and software development.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

HbA1c-Integrated Multimodal Deep Learning for Diabetic Retinopathy Severity Classification and Short-Term Progression Risk Stratification

Ananya V S[1], Jasmine Selvakumari Jeya I[2], Aldrin Joan Pandian W[3], Kannaiya Raja N[4], Indirani M[5], Saravanan S[6]

[1,3]School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Sehore, Madhya Pradesh, India, vsananya02@gmail.com, aldrinjoan6@gmail.com
[2,4]School of Computing Science and Engineering, VIT Bhopal University, Sehore, Madhya Pradesh, India, wjasminejeya@gmail.com, kannaiyaraja@vitbhopal.ac.in
[5]Department of Computer Science and, Business Systems, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India, indugovind10@gmail.com
[6]School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, India, jeyasaraa@gmail.com

Abstract.

Diabetic Retinopathy (DR) is a primary factor of reducible blindness among adults having diabetes, yet conventional fundus image screening provides only static severity grading and lacks quantitative estimation of progression risk or integration of key biomarkers such as glycated hemoglobin (HbA1c). This study presents a Multimodal DR-HbA1c framework, a lightweight deep learning architecture that performs five class DR severity classification from retinal fundus images while generating patient specific 6 and 12 month progression risk probabilities by combining predicted DR grade with a VTDR inspired rule based risk calculator calibrated from literature distributions. The model uses a pretrained DenseNet121 backbone producing 256 dimensional image features, a compact clinical branch encoding HbA1c into 16 dimensional features, and late fusion to form a 272 dimensional joint representation. The model attains 79.93% validation accuracy on 1,096 unseen APTOS 2019 images, enhancing 6.39% over the image only baseline with macro AUC 0.9185. Sensitivity calculation validates glycemic efficiency, exposing risk escalation with increasing HbA1c. Transparency applying LIME showcases clinically related retinal areas. This model proves the suitability of HbA1c determined personalized risk classification in DR screening for resource finite settings and helps adaptive screening periods for improved early intervention and management outputs.

Keywords. Multimodal Deep Learning, Progression Risk Prediction, Retinal Fundus Image Analysis, HbA1c, Diabetic Retinopathy (DR).

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

Diabetic Retinopathy is a leading microvascular is-sue of diabetes and the foremost effect of preventable vision loss in the employable age population worldwide. As of 2021, an estimated 103 million people globally were living with DR, with prevalence rising sharply in low- and middle-income nations-particularly India-due to the accelerating incidence of type 2 diabetes [1]. Current clinical screening methods depend solely on retinal fundus photo-graphs for static severity grading (No DR, Mild, Moderate, Severe, Proliferative). Although this method is effective for detection, it does not provide a quantitative assessment of the risk of future disease progression and does not con-sider essential clinical biomarkers, including glycated hemoglobin (HbA1c)[2]. To address this critical gap, the present work introduces the Multimodal DR-HbA1c framework -a novel lightweight deep-learning architecture that, performs five-class DR severity classification via a DenseNet121 image backbone and generates patient-specific 6- and 12-month progression risk probabilities through a VTDR-inspired clinical calculator that integrates the predicted DR grade with a single HbA1c input - enabling personalized risk stratification from the same retinal image under varying glycemic control [3]. A key clinical contribution is demonstrating that patients with identical retinal fundus appearances and thus identical DR grades from any image-only sys-tem can receive materially different 12-month progression risk estimates and screening interval recommendations depending on their glycemic control, a distinction that is entirely invisible to convention-al image-based screening [4]. The proposed system employs late-feature fusion of a frozen DenseNet121 image branch with a compact HbA1c clinical branch, a VTDR-inspired dual-horizon risk calculator, and LIME-based visual explainability all integrated into an end-to-end Gradio prototype. As the APTOS 2019 dataset contains no paired HbA1c labels, clinical HbA1c values were generated via literature-derived class-conditional normal distributions and used for both training and sensitivity analysis. The remainder of this paper presents the architecture, experimental results on 1,096 unseen validation images, and a sensitivity analysis demonstrating the clinical value of glycemic context in personalized DR risk stratification.

2. METHODOLOGY

The proposed Multimodal DR-HbA1c framework was developed to perform five-class DR severity classification from retinal fundus images and generate personalized 6- and 12-month progression risk probabilities by integrating the predicted grade with a HbA1c-aware clinical risk calculator. The image branch uses a pretrained DenseNet121 backbone (frozen) followed by global average pooling and a dense layer to extract 256-dimensional features. The clinical branch processes a single HbA1c value through two dense layers ($32 \rightarrow 16$ features). Late fusion concatenates both branches into a 272-dimensional representation, followed by dense layers ($128 \rightarrow 64$) and a 5-class softmax output. Only 306,517 parameters (4.2 % of total) are trainable. The model was trained on 2,566 retinal fundus illustration from the APTOS 2019 vision loss identification dataset, partitioned 70/30 into training and validation sets (2,566 train / 1,096 validation). The training class distribution was: No_DR (n=1,264, 49.3%), Moderate (n=700, 27.3%), Mild (n=259, 10.1%), Proliferative (n=207, 8.1%), and Severe (n=136, 5.3%). As the APTOS 2019 dataset contains no HbA1c labels, training and validation HbA1c values were synthetically generated using literature-derived class-conditional normal (No_DR: μ=6.8%, σ=1.2%; Mild: μ=7.8%, σ=1.4%; Moderate:

μ=8.6%, σ=1.5%; Severe: μ=9.8%, σ=1.3%; Proliferative: μ=10.2%, σ=1.4%), clipped to the clinical range [4.0%, 14.0%]. These parameters are derived from published HbA1c-DR correlations.Table 1 summarizes performance metrics across all models. The proposed multimodal model achieves the highest accuracy of 79.93 %, representing a +6.39 % improvement over the best single-image baseline (DenseNet121 at 73.54 %) and +4.84 % over the optimized ensemble. Macro F1-score is 0.5962 (driven by imbalance in Severe/Proliferative classes), macroaverage AUC is 0.9185, and micro-average AUC is 0.9589. Inference time is 27.6ms on Tesla T4 GPU.

Table 1. Efficiency analysis of baseline framework and Multimodal DR-HbA1c

Model	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Inference Time (ms)
ResNet50	66.42	0.421	0.392	0.377	16.0
ResNet18	69.43	0.463	0.429	0.424	6.1
MobileNetV2	69.53	0.509	0.536	0.512	25.9
DenseNet121	73.54	0.552	0.565	0.549	27.5
Simple Ensemble	75.00	0.574	0.556	0.559	~75.5
Optimized Ensemble	75.09	0.573	0.574	0.567	~75.5
Multimodal DR-HbA1c Framework	79.93	0.637	0.583	0.596	27.6

3. RESULT AND DISCUSSIONS

The incorporation of a single clinical biomarker-glycated hemoglobin (HbA1c)-with retinal fundus images meaningfully improves both classification accuracy and clinical utility. The proposed multimodal DR-HbA1c framework achieved 79.93% validation accuracy on the held-out validation set (n=1,096) of the APTOS 2019 dataset, surpassing the image-only DenseNet121 baseline by 6.39 percentage points while training only 4.2% of the total parameters. This gain demonstrates that HbA1c supplies complementary information beyond retinal texture alone. The most important clinical finding is that the same retinal images are linked to significantly different risks of disease progression and screening recommendations, depending on how well blood sugar is controlled. For example, a Mild DR retina shows 12-month development risk of 16.2% (requiring 6-monthly screening) at HbA1c 10.5%, evaluated with 9.9% (annual screening) at HbA1c 6.5%. This impact is sustained by a analytically important spearman correlation between HbA1c and 12 month risk ($\rho = 0.3706$, $p = 0.0438$, $n=30$) and strong within grade pearson associations ($r > 0.97$, $p < 0.01$) for four of five DR classes. Even retinas graded No_DR switch from low risk (9.9%, annual screening) under stable handling to moderate risk (16.2%, 6 monthly screening) under poor control (HbA1c 10.5%), illustrating a clinically practical adaptive by metabolic situation. Collectively, these results demonstrate the framework's capacity to move DR screening from static image-based grading toward personalized, actionable risk stratification [7].

4. CONCLUSIONS

Current DR screening primarily depends on retinal fundus images for static severity grading, offering no quantitative estimate of disease progression risk and overlooking key metabolic biomarkers such as HbA1c. This limitation results in uniform screening strategies that fail

to reflect individual patient risk profiles. To address this gap, a lightweight multi-modal deep learning framework was developed by combining a frozen pre-trained DenseNet121 image group with a compact clinical integrating HbA1c data through late-fusion combining. The approach is moreover improved with a VTDR-inspired logistic risk calculator and LIME-based clarity for enhanced interpretability. The suggested framework attained a validation accuracy (79.93%) on a held-out subset (n=1,096) of the APTOS 2019 dataset, illustrating a 6.39% enhancement over an image-only baseline, while preserving efficiency with 306,517 trainable features and a 27.6ms inference time. The equal retinal images produce notably distinct 12-month growth risks and screening suggestions over differing HbA1c levels from 6.5% to 10.5%, sustained by a analytically notable Spearman correlation (ρ = 0.3706, p = 0.0438). However, class imbalance affected recall for Severe and Proliferative DR categories. Since real HbA1c labels were unavailable, synthetic values were used and risk estimates were generated using a transparent rule-based approach rather than longitudinal data-driven modeling.

5. REFERENCES

[1] M. Kropp, O. Golubnitschaja, A. Mazurakova, L. Koklesova, N. Sargheini, T.K.S. Vo, E. de Clerck, J. Polivka Jr., P. Potuznik, J. Polivka, I. Stetkarova, P. Kubatka, G. Thumann, 'Diabetic retinopathy as the leading cause of blindness and early predictor of cascading complications—risks and mitigation', EPMA Journal, vol. 14, no. 1, pp. 21-42, 2023.

[2] J.K. Goh, C.Y. Cheung, S.S. Sim, P.C. Tan, G.S. Tan, T.Y. Wong, 'Retinal imaging techniques for diabetic retinopathy screening', Journal of Diabetes Science and Technology, vol. 10, no. 2, pp. 282-294, 2016.

[3] C.M.P. Jacoba, L.A. Celi, P.S. Silva, 'Biomarkers for progression in diabetic retinopathy: Expanding personalized medicine through integration of AI with electronic health records', Seminars in Ophthalmology, vol. 36, no. 4, pp. 250-257, 2021.

[4] Z. Li, L. Wang, X. Wu, J. Jiang, W. Qiang, H. Xie, H. Zhou, S. Wu, Y. Shao, W. Chen, 'Artificial intelligence in ophthalmology: The path to the real-world clinic', Cell Reports Medicine, vol. 4, no. 7, 2023.

[5] M.O. da Silva, A.E.C. do Carmo Chaves, G.C. Gobbato et al., 'Early choroidal and retinal changes detected by swept-source OCT in type 2 diabetes and their association with diabetic kidney disease: A longitudinal prospective study', BMC Ophthalmology, vol. 24, p. 85, 2024.

[6] S.L. Krishna, I.J.S. Jeya, S.N. Deepa, 'Fuzzy-twin proximal SVM kernel-based deep learning neural network model for hyperspectral image classification', Neural Computing & Applications, vol. 34, pp. 19343-19376.

[7] I.J. Selvakumari Jeya, S.N. Deepa, 'Lung cancer classification employing proposed real coded genetic algorithm based radial basis function neural network classifier', Computational and Mathematical Methods in Medicine, 2016, pp. 1-15.

Biographies



Ananya V. S. is an undergraduate student at the School of Computing Science, Engineering and Artificial Intelligence, VIT Bhopal University, Madhya Pradesh.



Jasmine Selvakumari Jeya I is currently a Senior Associate Professor with the School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh.



Aldrin Joan Pandian W is an undergraduate student at the School of Computing Science, Engineering and Artificial Intelligence, VIT Bhopal University, Madhya Pradesh.



Kannaiya Raja N is currently a Senior Associate Professor with the School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh.



Indirani M is currently an Associate Professor, Department of Computer Science and Business Systems, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, India.



Saravanan S is currently an Associate Professor, School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, India.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Trend-Aware Multi-Modal Generative Framework for Automated Social Media Content Creation

Devendra Veer[1], Smith Mavli[2], Sarthak Waykar[3], Dr. Chetan N. Aher[4]

[1,2,3,4] Department of Computer Engineering AISSMS Institute of Information Technology, Pune, India
[1]devendraveer0@gmail.com, [2]mavlismith84@gmail.com,
[3]sarthakwaykar7372@gmail.com, [4]chetan.aher07@gmail.com

Abstract.

Content development for social apps is step by step changing from disconnected, manual processes to lean, data-driven automated systems. After analyzing the initial trend data, the system creates proper text, logged campaign details in an organized database, and generates visuals like images and videos with coherent speech flow. These automated tools help social app creators find out the accurate time to post content and analyze how different people react to the published content, and fix a posting schedule which will work best for their engagement. Main motive of this method is to improve content creation, improved digital marketing processes and increase viewer engagement all while making sure dependable performance even when social trends change. The system is dependent on open-source models and cheap cloud services, so it will all together be cheaper and easy for generation.

Keywords. Generative AI, Social Media Automation, LLM, Multi-Modal Framework.

1. INTRODUCTION

Web content method has grown a lot lately, changing from just laboriously posting images and text to more upgraded systems that help create clever, common posting plans. Previously, handling social media mostly used simple scheduling tools that required regular manual checks and a lot of artistic work before any content was posted. These drawbacks have caused a demand for automated systems that can handle large amounts of trend data and generate digital content fast and smoothly [1].



Figure 1.1. System architecture of the AI-based Content Architect

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Latest developments in deep learning and generative AI have handled these problems by allowing exact and fast creation of text, images, and audio [3]. Scalability remains challenging when multiple accounts require simultaneous promotions, caption generation, and prompt processing. Existing social media tools often operate separately, focusing either on content creation or post scheduling. They also lack real-time trend adaptation and multimodal generation. These limitations increase workload and reduce efficiency. Therefore, this work proposes a unified, modular architecture integrating trend analysis, automated content generation, multimodal processing, and scheduling.

2. EXISTING TECHNOLOGY

Digital Marketing software was previously used for basic scheduling operations and for keeping track of analytics. Now, with things like Machine learning and Natural language processing, this system aids social media creators in gaining content engagement insights and creating content just by text prompts [4], [2]. Now, content services are divided into two main categories first is Social Media Management systems, which focuses on post scheduling and data analytics and second is Content Generation tools which generates text, images and videos.

2.1. Social Media Management (SMM) Systems

Posting on social media is an important function, but manual and fixed scheduling has proven to be quite faulty. Automation with AI can increase speed, but even now many lack the element of real-time adaptation to fast-changing trends, making them useless for today's creators.

2.2. Generative Creation Tools

Content creation now goes far beyond basic design tools. Adopting remote AI system, creators can create useful media files just by giving text steps [4], [5]. AI writing tools use Large Language Models to write captions, scripts for videos, and hashtags [4].

Table 1. Comparative analysis of content generation tools.

Metric	Traditional Methods	Generative Tools	Proposed System
Time Required	2–5 hours	30–60	< 3 minutes
Human Effort	High	Medium	Low
Automation Level	Low	Medium	High
Trend Adaptability	Low	Medium	High

Table 2. Performance metrics comparison (time, effort, and automation)

Content Type	Generation Latency	Human Editing Required	Accuracy
Text / Captions	< 15 Seconds	10%	94.8%
Images	< 40 Seconds	15%	92.1%

Synthesized Video	< 180 Seconds	20%	88.9%

These limitations stress the need of having a single system that involves content creation, trend analysis, and algorithmic publishing as one system.

3. EMERGING OPPORTUNITIES

As stated earlier, software systems have been used for timed publishing and creating bits and pieces of content in the digital marketing space. But with the coming growth of Content generation models and the availability of open-source Python libraries, there is much more scope and area of improvement in social media.

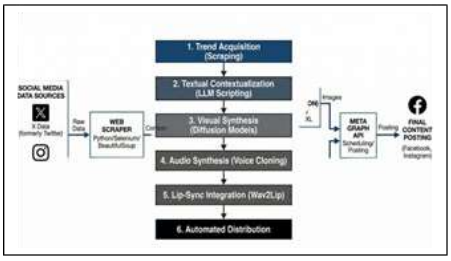


Figure 3.2. Integrated content generation pipeline creating image, audio, and video synthesis.

content creators system sequence. Over ongoing uses, intelligent automation can help Seamless channel integration, all platform support, which increases their value for digital creators.Brand security and contract is a big issue, particularly for approved content. Social media sites are deeply observed, heading to problems like copyright strikes or claims which can cause major cash flow issues. AI Content Analysis does routine checks like plagiarised content, checking copyrighted content or restricted topics [7].

4. METHODOLOGY

The AI Content Architect is built to transform online content which is trending into production ready content by using live data scrapping and machine learning techniques [5]. This approach follows a linear workflow that includes collection of trending data, writing prompts, generation of videos/text/images, and scheduling the output. This system improves constant posting schedule, notices how audience react to the posted content and quickly adapt to different social media sites.

4.1. System Workflow

The proposed pipeline automates content creation by collecting and preprocessing real-time social media trends. A large language model generates captions and scripts, which are transformed into images and videos. Content is scheduled through API-based automation tools, while engagement and reach metrics are monitored. This continuous workflow improves content quality, efficiency, scalability, and future performance.

4.2. Collecting and Processing Trend data

System collects unfiltered data from online sources like X and Instagram along with online available web scrapping tools. Data has all kinds of trending audio, texts, news and hashtags.

4.3. Content Generation Using LLMs

A large language model (LLM) acts as the primary tool for generating textual content [4], [5].Rather than older systems that used static structures, this model is designed to grasp situation better, allowing it to produce natural, relevant text.

4.4. Generating Visual Content

The content generated by the LLM is then used as input for an image creation tool, which uses a model like Stable Diffusion or DALL-E [6].

4.5. Scheduling and Analyzing Performance

This system uses cron jobs and the Meta Graph API to post content at the best times. After publishing, the system tracks how the content performs, such as likes and shares. These observations help social media creators rearrange their roadmap and get better at audience experience.

5. SYSTEM ARCHITECTURE

React.js and Next.js are used for creating an easy-to-navigate front-end, while Python and FastAPI serve the purpose of developing a scalable back-end. In addition, LangChain technology is employed for orchestrating LLMs in coordination with web scraping and PostgreSQL. There are APIs available for generating content in various formats including text, images, audio, and videos.

6. RESULTS AND DISCUSSION

Table 3. Evaluation of generation latency and accuracy.

Feature	**Legacy SMM Systems**	**Generative Tools**	**AI Architect**
Trend Detection	Manual Scrolling	None	Automated Scraping
Content Scope	Scheduling Only	Single-Modal	Multimodal
Platform Sync	API Limited	Disconnected	Unified Meta Graph

7. CONCLUSION

This article introduced a trend-based multi-modal content-creation system of automated social media. The system proposed combines real-time trend analysis, content generation and scheduling into a single pipeline. Through observed results, the system is observed to meaningfully reduce content generation time and lower manual labour but still produce high qualities. Future research will involve enhancing the quality of video generation, introducing real-time A/B testing and introducing adaptive learning processes to keep optimizing content performance.

8. REFERENCES

[1] J. Ahn and M. Kim, "Autonomous AI agents for multi-platform social media marketing: A simultaneous deployment study," Electronics, vol. 14, no. 21, p. 4161, Oct. 2025, doi: 10.3390/electronics14214161.

[2] T. Cetin Akinci and O. Topsakal, "Creating large language model applications utilizing LangChain," in Proc. Int. Conf. Appl. Eng. Nat. Sci. (ICAENS), Konya, Turkey, Jul. 2023, pp. 1050–1056.

[3] S. Beyari and A. Hashem, "Strategising with generative AI: Productivity gains in social media marketing," J. Marketing Commun., vol. 31, no. 4, pp. 412–430, Jan. 2025.

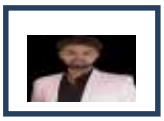
[4] R. K. P. Rakshitha et al., "Leveraging LLaMA3 and LangChain for rapid AI application development," J. Electrical Syst., vol. 20, no. 10s, pp. 2146–2153, 2024.

[5] P. Bansal and A. Ouda, "Study on integration of FastAPI and machine learning for continuous authentication of behavioral biometrics," in Proc. 2022 Int. Symp. Netw., Comput. and Commun. (ISNCC), 2022, pp. 1–6.

[6] R. D. Kumar, "Automating source checks in lab reports," in Proc. Int. Conf. Eng. Commun., Singapore, 2024, pp. 44–50.

[7] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," arXiv preprint arXiv:1312.6114, 2013.

Biographies



Devendra Veer is a Computer Engineering student at AISSMS Institute of Information Technology, Pune, India.



Smith Mavli is a Computer Engineering student at AISSMS Institute of Information Technology, Pune, India.



Sarthak Waykar is a Computer Engineering student at AISSMS Institute of Information Technology, Pune, India.



Dr. Chetan N. Aher is affiliated with the Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Toxicity Detection in online games using Reinforcement Learning

Mrs. Ruquia Taha Khatoon, Dantala Shivaghna, Katta Harsha Vardhan

[1]*Assistant Professor, Department of CSM, CMR College of Engineering & Technology, Medchal, Telangana, India.*
[2]*Student,CSE(AI&ML), CMR College of Engineering & Technology, Medchal, Telangana, India.*
[3]*Student,CSE(AI&ML), CMR College of Engineering & Technology, Medchal, Telangana, India.*

Abstract

Multiplayer online gaming provides entertainment and social interaction but can also encourage toxic chat behavior, including harassment and abuse. This paper presents an AI-based toxicity detection and moderation system integrated into a multiplayer Snake and Ladders game. Machine learning models trained on Kaggle datasets classify chat messages as toxic or non-toxic. A reinforcement learning agent learns appropriate moderation actions, such as issuing warnings or removing players, based on toxicity severity and frequency. The system integrates real-time chat functionality with a dynamic game board to support interactive gameplay. The proposed approach demonstrates how adaptive AI-driven moderation can promote safer, more enjoyable gaming environments while reducing online toxicity and supporting digital well-being.

Keywords. Toxicity Detection, Reinforcement Learning, Natural Language Processing, Real-Time Chat Moderation.

1. INTRODUCTION

People really like playing games on the internet. They play with others from places and ages. They work together. Make friends in these virtual worlds. They can make friends. They can have fun with people they would never meet in life [1]. There is a big problem, with online gaming. It is called toxicity. Online gaming has this problem. It is very serious [2]

Toxic behaviour in games is really bad [3-10]. It can be people harassing each other using hate speech swearing much bullying and being verbally mean [11]. Younger players and people who are already treated unfairly can be hurt the most by behaviour [12]. If we do not stop behaviour online games can become very unpleasant places to be. This can make players quit damage the trust between players and make online gaming not as welcoming as it should be to everyone. Toxic behaviour, in games is a big problem that needs to be addressed. The old way of moderating things, where people have to report problems or where we just filter out words is not working very well. This system is often slow. It does not work the same way every time. It just can't keep up with the number of things that are

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

happening in real time in modern games. Modern games have a lot of real-time interactions and this old system is not good enough to handle them.

Artificial Intelligence, Machine Learning, Natural Language Processing, and Reinforcement Learning enable intelligent systems to monitor player communication and behavior. These technologies analyze chat messages, identify toxic content, assess severity, and determine appropriate responses. Based on detected toxicity, the system can issue warnings, temporarily restrict communication, or remove players. This adaptive approach promotes respectful interaction, fairness, safety, and a positive multiplayer gaming environment.

2. RELATED WORK

In the gaming industry, there are various major platforms that have adopted AI-powered moderation tools. For example, companies such as Riot Games and Blizzard Entertainment use AI-powered chat filtering and behaviour analysis tools to monitor player interactions and analyze gamer behaviour in real time. However, most of the existing industrial solutions are heavily dependent on supervised learning approaches. Though supervised learning approaches are highly efficient in detecting explicit forms of abuse, they are often ineffective in dealing with slang, sarcasm, contextual insults, and dynamic patterns of toxic behavior. Moreover, most of these approaches are still based on traditional punishment schemes that are not adaptive to player behavior and context.

Recent studies have shown increased interest in applying Reinforcement Learning approaches for improving moderation approaches by enabling adaptive and sequential decision-making approaches.

Table 1 Comparison of Existing Toxicity Detection and Player Behaviour Analysis Techniques

Study	Data used	Key Techniques	Advantages	Limitations
Deep Learning of toxic comment classification	Kaggle Toxic comment dataset(Wikipedia comments)	Logistic Regression, CNN, LSTM	High classification accuracy, captures contextual patterns	Not specially designed for gaming chats, lacks real-time adaptation
Context-Aware Toxicity Detection Using BERT	Social media and gaming chats datasets	BERT, Transformer-based NLP models	Better understanding of context, sarcasm and subtle toxicity	Computationally expensive, requires large training data
Real-time Chat Moderation in online games	n-game chat logs from multiplayer games	Keyword filtering + Supervised ML classifier	Fast detection, scalable for large platforms	Rule-based filtering struggles with evolving slang and coded language

Behaviour - based Toxicity player detection (Riot Games research)	Player chat logs + behavioral reports	NLP + Player behaviour scoring models	Combines chat analysis with player history, improves fairness	Static punishment policies, limited adaptive decision-making

3. SYSTEM ARCHITECTURE

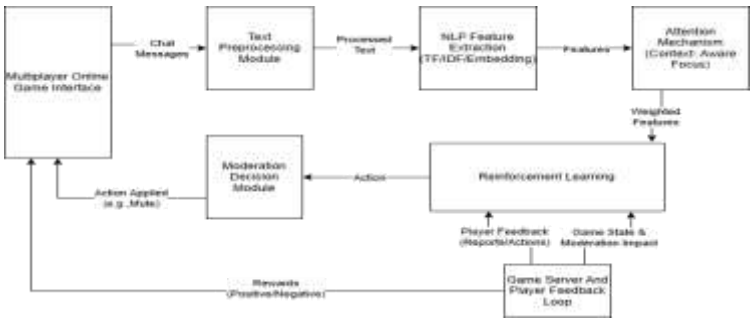


Fig1: System Architecture of Toxicity Detection in Online Games Using Reinforcement Learning

4. FLOWCHART

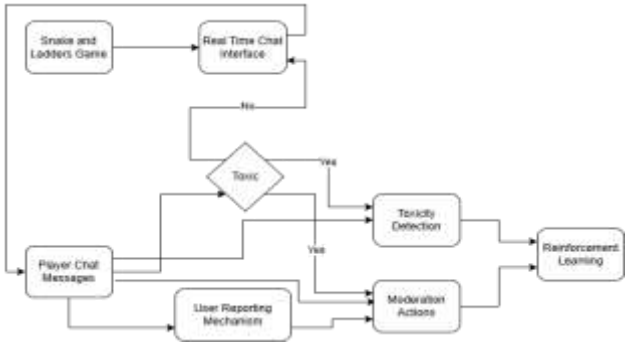


Fig2: Flow Chart of Toxicity detection in Online games using Reinforcement Learning

5. REFERENCES

[1] Rosenfeld, A., & Richardson, A. (2021). The Role of Artificial Intelligence in Detecting and Preventing Online Toxicity. IEEE Transactions on Games.

[2] Papangelis, A., et al. (2020). Toxicity Detection in Online Games: A Context-Aware NLP Approach. Proceedings of the IEEE Conference on Games (CoG).

[3] Blackburn, J., Kwak, H., & Lee, S. (2022). STFU Noob!: Predicting Crowdsourced Decisions on Toxic Behavior in Online Games. ACM Conference on World Wide Web.

[4] S. Gore and C. Puthillate, "Authentication and authorization of users in an information handling system between baseboard management controller and host operating system users," U.S. Patent 11,038,874, 2021.

[5] Kaggle Dataset. uetchy/sensai – Toxic Comment Classification Dataset. https://www.kaggle.com/datasets/uetchy/sensai.

[6] Gao, Y., et al. (2022). Real-Time Toxicity Detection and Intervention in Online Games Using Reinforcement Learning. Journal of AI Research & Applications.

[7] V. Antony Asir Daniel, K. Vijayalakshmi, Priyanka Pramod Pawar, Deepak Kumar, A. Bhuvanesh, A. Josephine Christilda,Enhanced affinity propagation clustering with a modified extreme learning machine for segmentation and classification of hyperspectral imaging,e-Prime - Advances in Electrical Engineering, Electronics and Energy, Volume 9, 2024, 100704, ISSN 2772-6711, https://doi.org/10.1016/j.prime.2024.100704.

[8] Y. Yang, J. Li, Z. Wang, and H. Zhang,"Towards detecting contextual real-time toxicity for in-game chat," Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023.

[9] Tereshchenko and M. Hämäläinen,"Efficient toxicity detection in gaming chats: A comparative study," arXiv Preprint, 2025.

[10] Cornel, J., et al. (2025). "Comparative Analysis of Two-Class and Multi-Class Toxicity Detection Using Multi-Source Gaming Chat Data.".

[11] T. Fesalbon, L. Perez, and J. Martin,"Fine-tuning pre-trained language models to detect in-game trash talks," arXiv Preprint, 2024.

[12] U. Naseem, M. Imran, and R. Razzak,"GameTox: A comprehensive dataset and analysis for enhanced toxicity detection in online games," Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL), 2025.

Biographies



Ruquia Taha Khatoon is currently working as an Assistant Professor in the Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning), CMR College of Engineering & Technology, Medchal, Telangana, India. Her areas of interest include Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, and Data Science. She is actively involved in teaching, academic research, and mentoring students in emerging technologies. Her research interests focus on the development and application of intelligent computing techniques to solve real-world problems.



Dantala Shivaghna received her Bachelor of Technology (B.Tech) degree in Computer Science and Engineering (Artificial Intelligence & Machine Learning) from CMR College of Engineering & Technology, Medchal, Telangana, India. Her research interests include Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, and Data Science. She is interested in exploring emerging technologies and their applications in real-world problems. She has been actively involved in academic projects and research activities in the field of Artificial Intelligence and Machine Learning.



Katta Harsha Vardhan received her Bachelor of Technology (B.Tech) degree in Computer Science and Engineering (Artificial Intelligence & Machine Learning) from CMR College of Engineering & Technology, Medchal, Telangana, India. Her research interests include Artificial Intelligence, Machine Learning, Deep Learning, Natural Language Processing, and Data Science. She is interested in exploring emerging technologies and their applications in real-world problems. She has been actively involved in academic projects and research activities in the field of Artificial Intelligence and Machine Learning.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

MedVerify: A Severity-Aware System for Verification of Medical Claims Generated by Conversational AI Systems

Aditi Londhe[1], Aarya Badhe[2], Tanish Charthankar[3], Shrujan Kundgulwar[4], Sachin Shelke[5]

[1,2,3,4,5]SCTR's Pune Institute of Computer Technology, Department of Information Technology, Pune, India
aditivlondhe@gmail.com[1], ryabadhe11@gmail.com[2], vctanish7@gmail.com[3], shrujankundgulwar47@gmail.com[4], sachindshelke@pict.edu[5]

Abstract.

Large Language Models (LLMs) are increasingly used for health-related queries but may generate hallucinated or medically unsupported information, creating significant risks for patient safety. This paper presents MedVerify, a severity-aware verification tool that identifies, prioritizes, retrieves, and validates medical claims generated by conversational AI. The system combines structured drug-fact verification using FDA drug-label data with evidence-based Natural Language Inference using biomedical text. Implemented as a browser extension, MedVerify verifies chatbot responses against trusted biomedical databases and provides claim-level verification results, severity indicators, and supporting references. This approach enables transparent, evidence-based validation of AI-generated medical information in real time.

Keywords. Medical Claim Verification, Conversational AI, Large Language Models, Medical Hallucination, Severity-Aware Verification.

1. INTRODUCTION

Large Language Models (LLMs) are increasingly used to obtain medical information, increasing reliance on conversational AI in healthcare. Users frequently seek information about symptoms, medications, treatments, and health conditions. Although LLMs can provide fluent and informative responses, they may generate hallucinated content that appears factual but lacks reliable evidence or cannot be verified. Such hallucinations are particularly concerning in healthcare because incorrect dosage instructions, fabricated drug interactions, or unsupported treatment recommendations may cause serious harm. Existing research on hallucination detection and fact-checking primarily focuses on general-domain information rather than medical knowledge. Medical verification is more challenging because biomedical information requires structured knowledge, authoritative evidence, and high reliability. To address this gap, this paper presents MedVerify, a browser-based medical claim verification tool. MedVerify analyzes conversational AI outputs, identifies medical claims, retrieves supporting evidence from trusted biomedical databases, and validates

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

claims using evidence-based verification methods. The system provides users with reliable, transparent, and evidence-supported verification of AI-generated medical information.

2. RELATED WORK

Research in hallucinations has increased in natural language generation (NLG). Ji et al. [6] gave an extensive review and taxonomy of hallucinations in NLG and created a benchmark for future research in hallucination detection. TruthfulQA, developed by Lin et al. [9] is a dataset that helps in assessing the ability of language models to faithfully emulate human lies and showed that bigger models are no more truthful than smaller models. In the field of medicine, Chelli et al. [4] performed a systematic analysis of hallucinations in medical LLMs.

A Retrieval-Augmented Generation (RAG) model was proposed by Lewis et al. [8], which verifies the generated outputs using evidence retrieval and represents a vital constituent of the evidence retrieval module in MedVerify. There exist pretrained models for biomedical language understanding in the form of BioBERT by Lee et al. [7], PubMedBERT by Gu et al. [5], and BioMedBERT by Chakraborty et al. [3]. In the context of entailment relations for medical QA systems by Abacha and Demner-Fushman [1]. Alsentzer et al. [2] have come up with clinical BERT embeddings for biomedical purposes. Luo et al. [10].

3. KEY CONTRIBUTION

MedVerify introduces risk-based medical claim classification, combines structured drug verification with evidence-based Natural Language Inference, and provides real-time chatbot response verification through a Chrome Extension without modifying underlying AI models.

4. METHOD, EXPERIMENTS AND RESULTS

4.1. System Architecture

The verification pipeline of MedVerify is modular and multi-stage. First, when MedVerify receives an AI response (R), it collects a list of all of the potential claims that were made in that response.

$$C = \{c_1, c_2, \ldots, c_n\} \quad (1)$$

Next, all potential claims (C) are run through a sequential process as shown in Table 1:

$$R \rightarrow C \rightarrow S(C) \rightarrow E(C) \rightarrow V(C) \rightarrow Score(C) \quad (2)$$

where:
C = Extracted Claims, S(C) = Severity Classification, E(C) = Retrieved Evidence, V(C) = Verification Outcome, Score(C) = Final Confidence Score.

4.2. Medical Claim Extraction

The primary objective of this stage is to determine whether there are any medical claims (which may include references to medication, suggestions of beneficial results from medications, dosage descriptions, treatment suggestions, side effect disclosures, diagnostic procedures and interactions between medications) in the AI-derived response.

4.3. Severity Classification

MedVerify categorizes every claim c*i* with one of four distinct severity labels:

$$S(c_i) \in \{Critical, Serious, Moderate, Low\} \quad (3)$$

The categorization is made using a rules-based classification technique which relies on three primary factors to classify whether a claim should be assigned a severity label: (i) dosage expressions (e.g. mg, frequency) present (ii) drug-related entities such as medications, interactions (iii) clinical action verbs such as prescribe, diagnose and treat. Severity weights are assigned to each of the four severity levels to be used when calculating the total score for each claim as follows:

$$SeverityWeight(c) = 1.5\ (Critical), 1.2\ (Serious), 1.0\ (Moderate), 0.8\ (Low) \quad (4)$$

4.4. Evidence Retrieval

The evidence retrieval component finds trustworthy biomedical literature (peer-reviewed), trustworthy labels & safety information for medications (OpenFDA), trustworthy consumer health information (MedlinePlus), and Wikipedia (general context and fallback information) to support or not support statements made by individual claims. Each claim c has an associated query that is generated using the keywords associated with the entities connected with the claim:

$$Q(c) = keywords(entities(c)) \quad (5)$$

The system obtains the top k document(s) (k = 3–5) whose keyword relevance score is the highest and has the most shared entity types and then pulls out the relevant sentence(s) needed to verify that individual claim(s) c are true:

$$E(c) = \{e_1, e_2, \ldots, e_k\} \quad (6)$$

4.5. Hybrid Verification Framework

The MedVerify Tool uses a Hybrid Verification framework which includes two ways to verify claims:Structured Drug Fact Verification: Uses OpenFDA databases for verification against FDA approved drug label dosage, contraindications, & interaction data structured in the database, allowing for precise comparison of the attributes at the attribute level.Evidence-Based NLI Verification: Utilizes a biomedical NLI to verify general medical claims retrieved from electronic sources such that each claim is classified as either Entailment, Contradiction, or Neutral. Any claim that has been classified as a contradiction will be flagged as a disputed claim.

4.6. Medical Confidence Scoring

Verification results, source authority, and severity weight are combined into a final confidence score.

$$\text{Score(c)} = \text{V(c)} \times \text{Authority(e)} \times \text{SeverityWeight(c)} \quad (7)$$

Claims lacking sufficient evidence are labelled as Cannot Verify.

4.7. System Implementation

MedVerify, built as a Manifest V3 Chrome extension, monitors chatbot responses using PubMed, OpenFDA, and MedlinePlus APIs for evidence-based verification.

4.8. Evaluation

MedVerify analysed 100 medical claims, identifying various hallucination severities and verifying supporting evidence within an average response time of 2.3 seconds.

Table 1. Performance of MedVerify.

Category	Precision	Recall	F1	Samples
Critical	0.91	0.88	0.89	25
Serious	0.87	0.85	0.86	25
Moderate	0.82	0.80	0.81	25
Low	0.78	0.76	0.77	25
Overall	0.85	0.82	0.83	100

5. DISCUSSION

MedVerify is most accurate for claims of critical severity, while moderate and low-severity claims tend to be less accurate owing to complicated medical data. The limitations involve reliance on medical evidence that may be available, limited coverage of rare illnesses and recently approved medicines, and complications with understanding medical language. MedVerify facilitates medical decision-making but cannot substitute professional doctors.

6. CONCLUSION

MedVerify integrates biomedical NER, severity classification, multi-source evidence retrieval, and hybrid NLI-based verification. Testing on 100 claims achieved an overall F1-score of 0.83, with 0.89 for critical claims and 2.3-second processing. Future work includes advanced NLI models, broader databases, multilingual support, and user studies.

7. REFERENCES

[1] Abacha, S.B. & Demner-Fushman, D. 2020. A question-entailment approach to question answering. *CEUR Workshop Proc.* 2696: 1–10.

[2] Alsentzer, E., Murphy, J., Boag, W., Weng, W.-H., Jin, D., Naumann, T. & McDermott, M. 2019. Publicly available clinical BERT embeddings. In *Proc. Workshop Biomed. Natural Language Process. (BioNLP)*, Florence, Italy: 72–78.

[3] Chakraborty, S., Saha, S. & Bandyopadhyay, S. 2020. BioMedBERT: A pre-trained biomedical language model for question answering and information retrieval. In *Proc. Int. Conf. Comput. Linguistics (COLING)*, Barcelona, Spain: 669–679.

[4] Chelli, M., Descamps, J., Lavoué, A., Trojani, M., Azar, P., Deckert, J., Raynier, A., Clowez, J.-B., Kün, S. & Ruetsch-Chelli, A. 2024. Hallucination in medical LLMs: A systematic evaluation. *JMIR Med. Informatics* 12(1): e53164.

[5] Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J. & Poon, H. 2022. Domain-specific language model pretraining for biomedical natural language processing. *ACM Trans. Comput. Healthcare* 3(1): 1–23.

[6] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A. & Fung, P. 2023. Survey of hallucination in natural language generation. *ACM Comput. Surv.* 55(12): 1–38.

[7] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H. & Kang, J. 2020. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36(4): 1234–1240.

[8] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. & Kiela, D. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, Virtual Event, vol. 33: 9459–9474.

[9] Lin, S., Hilton, J. & Evans, O. 2022. TruthfulQA: Measuring how models mimic human falsehoods. In *Proc. Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Dublin, Ireland: 3214–3252.

[10] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H. & Liu, T.-Y. 2022. BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Briefings Bioinformatics* 23(6): 1–10.

Biographies



Aditi Londhe. Student, SCTR's Pune Institute of Computer Technology, Department of Information Technology, Pune, India.



Aarya Badhe. Student, SCTR's Pune Institute of Computer Technology, Department of Information Technology, Pune, India.



Tanish Charthankar. Student, SCTR's Pune Institute of Computer Technology, Department of Information Technology, Pune, India.



Shrujan Kundgulwar. Student, SCTR's Pune Institute of Computer Technology, Department of Information Technology, Pune, India



Sachin Shelke. Assistant Professor, SCTR's Pune Institute of Computer Technology, Department of Information Technology, Pune, India.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Blockchain-Integrated and Privacy-Preserving AI Framework for Drug Authenticity Verification in Intelligent Pharmaceutical Supply Chains

Shamali A. Gupte[1] and Rajkumar Shende[2]

[1,2]Department of Computer Engineering, St. Francis Institute of Technology (SFIT), Mumbai, India
[1]shamaligupte@student.sfit.ac.in, [2]rajkumarshende@sfit.ac.in

Abstract.

Fake medications pose serious risks to patient safety and healthcare trust, particularly in developing nations. Conventional authentication methods can verify product identifiers but remain vulnerable to duplication. This paper introduces DAVASC, a decentralized authenticity verification architecture integrating permissioned Hyperledger Fabric, a GAT-Transformer model, and Secure Multi-Party Computation (SMPC). A Merkle-DAG mechanism securely anchors transaction records across permissioned and public ledgers while supporting regulatory audits. Evaluated on 120,000 synthetic pharmaceutical transactions covering six counterfeit injection methods, DAVASC achieves 97.4% centralized and 96.8% SMPC-based detection accuracy, 68.3 TPS throughput, and confirmation times below 12 milliseconds, demonstrating secure, scalable, and reliable counterfeit detection.

Keywords: Pharmaceutical Supply Chain Security; Drug Authenticity Verification; Hyperledger Fabric; Graph Attention Network; Secure Multi-Party Computation; Merkle-DAG.

1. INTRODUCTION

The global pharmaceutical market reached an estimated value of USD 1.48 trillion in 2022, yet a substantial share of medicines in developing nations still fail to meet quality and authenticity standards. The World Health Organization (WHO) estimates that about 10% of medicines worldwide are substandard or falsified [1]. The 2009 heparin contamination incident, which caused nearly 250 fatalities in the United States, and the ongoing theft of antimalarial drugs across sub-Saharan Africa illustrate how supply chain failures translate directly into clinical harm. The US Drug Supply Chain Security Act (DSCSA), the EU Falsified Medicines Directive (FMD), and WHO prequalification programs have introduced stronger traceability requirements, but markets most prone to counterfeiting still lack the infrastructure to comply with them. Prior work has explored IoT-based cold-chain monitoring, cross-border consortium traceability, zero-knowledge price attestation, and univariate anomaly detection on transaction logs, but none of these approaches combine relational graph modeling with SMPC-level privacy.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Blockchain technology provides a shared, protected ledger that multiple parties can access to record custody activity with permanent integrity. Smart contracts extend this by automating enforcement actions without relying on a centralized authority. This study applies graph-based deep learning to identify genuine anomalies within high-volume transaction streams. By combining a permissioned Hyperledger Fabric ledger with a GAT-T detection model trained under SMPC, DAVASC delivers regulatory-grade auditability, commercial confidentiality, and real-time counterfeit detection.

1.1. Principal Contributions

The principal contributions of this work are as follows:DAVASC integrates Hyperledger Fabric, privacy-preserving GAT-T detection, Merkle-DAG anchoring, and SMPC training to secure pharmaceutical supply chains, achieving 97.4% accuracy, 96.8% SMPC accuracy, and 68.3 TPS.

2. RELATED WORK

Blockchain adoption in pharmaceutical supply chains has largely focused on traceability rather than behavioral intelligence. Uddin [6] developed permissioned Hyperledger Fabric ledger designs for drug traceability, and Bandhu et al. [7] demonstrated smart contract-based recall mechanisms, but neither incorporated a fraud detection layer. Singh et al. [2] addressed IoT-based cold-chain monitoring, and Musamih et al. [3] addressed cross-border drug traceability, while Uddin et al. [4] introduced blockchain architectures with zero-knowledge attestation, a confidentiality mechanism that DAVASC extends through SMPC. On the machine learning side, Alqarni et al. [5] explored smart contract automation in supply chains; Wang et al. [8] showed that graph convolutional networks are effective fraud detectors in relational transaction data; Naga Sudha and Nayahi [9] showed that Hyperledger-based temporal supply chain modeling captures multi-transaction patterns; and Padma and Ramaiah [10] examined secure information sharing in pharmaceutical supply chains using blockchain and smart contracts. No prior system combines graph-structured relational modeling, SMPC-level privacy, cross-chain auditability, and automated enforcement in a single framework.

3. PROPOSED DAVASC FRAMEWORK

3.1. Hyperledger Fabric Design with Merkle-DAG Anchoring

DAVASC runs on Hyperledger Fabric version 2.5 with four ordering-service nodes operating under Raft consensus and eight peer nodes distributed across manufacturer, distributor, wholesaler, and pharmacy organizations.The Merkle-DAG anchoring mechanism works as follows. Every Δ_t = 500 transactions, the DAVASC anchor service computes a Merkle root M_n from the transaction hashes in the current batch interval:

$$M_{\mathrm{n}} = MerkleRoot(\, H(tx_{\mathrm{i}}) : i \in [n \cdot \Delta t, (n+1) \cdot \Delta t)\,) \quad (1)$$

3.2. GAT-T Anomaly Detection Model

The Graph Attention Network component of GAT-T computes, for each transaction edge e = (u, v, t), a learned representation of how structurally unusual that transaction is relative to

its local network context. For node u with neighbor set N(u), the attention coefficient α_{ij} is defined as:

$$\alpha_{ij} = softmax_i(LeakyReLU(a^T \cdot [W h_i || W h_j])) \quad (2)$$

GAT uses multi-head attention to model relationships within the supply chain. The LSTM-enhanced Transformer, on the other hand, is used for analyzing transaction sequences and recognizing counterfeit activity patterns:

$$A(e) = \sigma(W_2 \cdot ReLU(W_1 \cdot [zGAT || zTrans] + b_1) + b_2) \quad (3)$$

When A(e) exceeds the threshold $\tau = 0.78$, the smart contract freezes the affected batch, raises a tamper-evident alert on the public chain, and sends HTTPS notifications to the relevant regulatory authority and all downstream holders of the asset.

3.3. SMPC Training Protocol

DAVASC trains the GAT-T model using SPDZ, a multi-party computation protocol that provides strong cryptographic security against a dishonest majority of participants under a semi-honest threat model. Each stakeholder node i holds a secret share $[w]_i$ of the global model weight vector w.

$$\Pi SMPC : \forall i,j \in [K], i \neq j : I([w]_i, D_j) = 0 \quad (4)$$

This expression states that the mutual information between node i's model share and node j's dataset is zero: observing node i's share reveals nothing about node j's data.

4. EXPERIMENTAL METHODOLOGY

4.1. Dataset

DAVASC was evaluated on a synthetic dataset of 120,000 pharmaceutical transactions generated across 12 WHO-modeled supply chain topologies spanning January 2020 to December 2023. Each record includes a batch identifier, drug classification, custody parties, timestamp, geolocation hash, price deviation, temperature flag, and certification status. Six counterfeit strategies (serial cloning, price undercutting, routing bypass, certification reuse, temperature falsification, and multi-hop laundering) were embedded at an overall prevalence of 9.1%, with the data split 70/15/15 for training, validation, and testing.

4.2. Baselines and Evaluation Metrics

Five baseline models were compared against the proposed GAT-T architecture: Isolation Forest, Random Forest, a Support Vector Machine with an RBF kernel, a standalone Graph Convolutional Network (GCN), and a standalone Transformer. All models were evaluated on Accuracy, Precision, Recall, and F1-Score using a held-out test set of 18,000 samples at 9.1% counterfeit prevalence. Blockchain performance was benchmarked against a standard Hyperledger Fabric (v2.5) baseline and a Hyperledger Besu private chain.

5. RESULTS AND DISCUSSION

5.1. Machine Learning Model Performance

Table 1 shows GAT-T achieving the highest performance, with 97.4% accuracy and 97.1% F1-score. SMPC training achieves 96.8% accuracy, demonstrating minimal privacy-related performance loss.

Table 1. Classification performance comparison across ML models on the test set.

ML Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Isolation Forest	88.4	87.9	86.1	87.0
Random Forest	90.7	89.2	91.1	90.1
SVM (RBF Kernel)	87.2	85.8	84.6	85.2
Standalone GCN	93.1	92.5	91.8	92.1
Standalone Transformer	92.7	91.3	92.9	92.1
GAT-T Centralized	97.4	96.9	97.3	97.1
GAT-T SMPC (Proposed)	96.8	96.2	96.7	96.4

5.2. Blockchain and System Performance

Table 2 reports blockchain performance metrics for DAVASC against a standard Hyperledger Fabric baseline and a Hyperledger Besu private chain configuration comparable to those used in prior pharmaceutical traceability work.

Table 2. DAVASC vs. Hyperledger Fabric baseline and Hyperledger Besu.

Metric	H. Fabric Baseline	H. Besu Private	DAVASC (Proposed)	Unit
Throughput	40-50	35-45	68.3	TPS
Avg. Block Latency	~500	~800	< 12	ms
Smart Contract Exec.	~320	~450	~140	ms
Energy/Transaction	~0.008	~0.012	~0.003	J
Privacy Mechanism	Channel isolation	None	SMPC + Merkle-DAG	—
Cross-chain Audit	No	No	Yes (Ethereum anchor)	—
Attack Resistance	Partial	Partial	Full (4 vectors)	—

6. CONCLUSION AND FUTURE WORK

This paper introduced DAVASC, a decentralized framework combining Hyperledger Fabric, GAT-Transformer anomaly detection, SMPC-based privacy, and Merkle-DAG anchoring for pharmaceutical supply-chain security. The system achieved 97.4% centralized accuracy, 97.1% F1-score, and 96.8% accuracy under SMPC, with only 0.6 percentage-point privacy loss. Hyperledger Fabric delivered 68.3 TPS with sub-12 ms latency, while attack detection ranged from 87% to 99.2%. Future work includes hospital-pharmacy deployment, asynchronous SMPC, and GDPR-compliant chaincode.

7. REFERENCES

[1] World Health Organization, "Substandard and falsified medical products," WHO Fact Sheet, 2023. https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products.

[2] R. Singh, A. D. Dwivedi, and G. Srivastava, "Internet of Things based blockchain for temperature monitoring and counterfeit pharmaceutical prevention," Sensors, vol. 20, no. 14, p. 3951, 2020. https://doi.org/10.3390/s20143951.

[3] A. Musamih, K. Salah, R. Jayaraman, J. Arshad, M. Debe, Y. Al-Hammadi, and S. Ellahham, "A blockchain-based approach for drug traceability in healthcare supply chain," IEEE Access, vol. 9, pp. 9728–9743, 2021. https://doi.org/10.1109/ACCESS.2021.3049920.

[4] M. Uddin, K. Salah, R. Jayaraman, S. Pesic, and S. Ellahham, "Blockchain for drug traceability: Architectures and open challenges," Health Informatics Journal, vol. 27, no. 2, p. 14604582211011228, 2021. https://doi.org/10.1177/14604582211011228.

[5] M. A. Alqarni, M. S. Alkatheiri, S. H. Chauhdary, and S. Saleem, "Use of blockchain-based smart contracts in logistics and supply chains," Electronics, vol. 12, no. 6, p. 1340, 2023. https://doi.org/10.3390/electronics12061340.

[6] M. Uddin, "Blockchain Medledger: Hyperledger Fabric enabled drug traceability system for counterfeit drugs in pharmaceutical industry," International Journal of Pharmaceutics, vol. 597, p. 120235, 2021. https://doi.org/10.1016/j.ijpharm.2021.120235.

[7] K. C. Bandhu, R. Litoriya, P. Lowanshi, M. Jindal, L. Chouhan, and S. Jain, "Making drug supply chain secure traceable and efficient: a blockchain and smart contract based implementation," Multimedia Tools and Applications, vol. 82, no. 15, pp. 23541–23568, 2023. https://doi.org/10.1007/s11042-022-14238-4.

[8] C. Wang, M. Wang, X. Wang, et al., "Multi-relational graph representation learning for financial statement fraud detection," Big Data Mining and Analytics, vol. 7, no. 3, pp. 920–941, 2024. https://doi.org/10.26599/BDMA.2024.9020013.

[9] C. M. Naga Sudha and J. V. Nayahi, "TrackChain: Hyperledger based pharmaceutical supply chain — resource utilization perspective," Heliyon, vol. 10, no. 1, p. e23250, 2024. https://doi.org/10.1016/j.heliyon.2023.e23250.

[10] A. Padma and M. Ramaiah, "Blockchain based solution for secure information sharing in pharma supply chain management," Heliyon, vol. 10, no. 22, p. e40273, 2024. https://doi.org/10.1016/j.heliyon.2024.e40273.

Biographies



Shamali A. Gupte is a student in the Department of Computer Engineering at St. Francis Institute of Technology (SFIT), Mumbai.



Rajkumar Shende is a faculty member in the Department of Computer Engineering at St. Francis Institute of Technology (SFIT), Mumbai.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

AI-Driven Parcel Sorting System Using QR Code Recognition and PLC Control

Dr. G. S. Gadge[1], Rushikesh Kokate[2], Aditya Jamdade[3], Kunal Danawale[4]

[1,2,3,4] Department of Electrical Engineering, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati, India. gaurav.gadge@vpkbiet.org, kokaterushikesh86@gmail.com, jamdadea196@gmail.com, kunaldanawale9696@gmail.com

Abstract.

The increase in parcel traffic caused by e-commerce and courier services has created a need for quicker and more reliable sorting methods. In many small and medium facilities, parcel sorting is still performed manually, which can lead to delays and routing mistakes. This paper presents a parcel sorting setup that uses QR code recognition with PLC-based control. Parcels move on a conveyor belt, and a webcam captures QR labels during movement. OpenCV is used to read the QR data, which is then sent through an Arduino Mega 2560 to a T-950 PLC. According to the decoded destination, the PLC activates pneumatic cylinders that push parcels into separate bins. Trials were carried out using 50 parcels for three destinations. The system achieved an overall sorting accuracy of 98% with an average response time below 200 ms. The developed setup can be useful in courier offices, warehouses, and training laboratories.

Keywords. Parcel Sorting; QR Code Recognition; PLC Control; OpenCV; Pneumatic Automation

1. INTRODUCTION

The logistics sector in India is gradually moving toward automation because of the growing number of e-commerce deliveries. In many smaller warehouses, parcel sorting is still done manually, which can be slow and may lead to routing mistakes. Manual handling also becomes difficult when parcel volume increases during busy periods. These issues create a need for simple and affordable automation systems.To address this problem, a low-cost parcel sorting system was developed using computer vision and PLC control. A standard HD webcam captures QR labels, while Python with OpenCV reads the destination details in real time [1]. An Arduino Mega 2560 is used as a communication link between the computer and the T-950 PLC. Based on the decoded destination, the PLC operates pneumatic actuators to direct parcels into separate bins.This setup provides a practical and economical alternative to expensive commercial sorting systems [3], [4]. It can be useful for courier offices, warehouse laboratories, and small to medium logistics facilities.

2. RELATED WORK

Recent studies indicate that computer vision is playing an important role in logistics and warehouse automation. Vision-based systems are widely used for package recognition,

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

inspection, tracking, and sorting, helping industries reduce manual effort and improve operational efficiency [1], [6].

Image processing techniques using cameras and OpenCV have also been applied for parcel identification, barcode reading, and QR label detection in logistics environments. These methods are suitable for real-time applications because they are cost-effective, flexible, and easier to implement than complex industrial vision systems [1].PLC-based conveyor systems are commonly used in industries for automatic material handling and sorting because of their reliability, fast response, and simple control structure [3]. Pneumatic actuator systems controlled through PLCs are also widely used to divert products into separate bins or lanes in automated processes [4].Recent work has additionally focused on parcel measurement, intelligent package recognition, and QR-based data registration to improve warehouse tracking and sorting performance [5], [7], [8]. Compared with earlier approaches, the present work combines QR code recognition using OpenCV, PLC control, and pneumatic sorting in one compact and low-cost setup suitable for small and medium logistics applications.

3. METHOD, EXPERIMENTS AND RESULTS

The developed system consists of a conveyor belt, webcam, personal computer, Arduino Mega 2560, relay module, T-950 PLC, solenoid valves, pneumatic cylinders, and sorting bins.

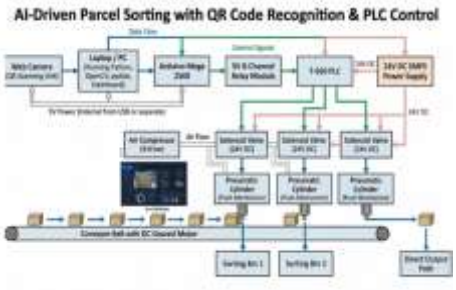


Figure 3.1. AI-Driven Parcel Sorting System Architecture (Author Developed)



Figure 3.2. Pneumatic Cylinder-Based Sorting Mechanism



Figure 3.3. T-950 PLC Controller

System performance was evaluated across 50 test parcels with QR codes for three destinations: Pune, Baramati, and Indapur. Table 2 summarizes the results, and Figure 6 shows the accuracy comparison.

During physical trials conducted at the VP Center of Excellence in Digital Manufacturing & Robotics, the system successfully identified and sorted all parcels destined for Pune and Indapur. One sorting error occurred in the Baramati batch because the QR label was partially obscured by the parcel packaging, preventing clear image capture by the webcam. The response time remained below 200 ms between QR detection and activation of the T-950 PLC-controlled pneumatic arm.

Table 2. Performance Evaluation of Parcel Sorting System

Destination	Total Parcels	Correctly Sorted	Accuracy (%)
Pune	20	20	100
Baramati	15	14	93.3
Indapur	15	15	100
Overall	**50**	**49**	**98%**



Figure 3.4. Experimental Setup



Figure 5. Parcel Sorting Dashboard

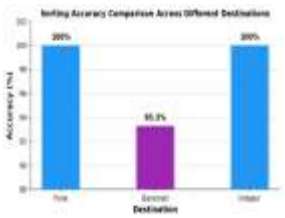


Figure 6. Accuracy Chart

4. DISCUSSIONS

During testing, the system operated reliably and maintained accurate sorting performance. All parcels for Pune and Indapur were sorted correctly, while one parcel in the Baramati batch was misrouted because the QR label was partially hidden during scanning. The average response time from detection to actuation remained below 200 ms, showing that the setup is suitable for small and medium parcel handling applications. Compared with manual sorting, the developed system reduces labor effort, improves consistency, and increases processing speed.

5. CONCLUSIONS

Motivation: This work was driven by the need to reduce the inefficiencies and high error rates associated with manual parcel sorting in small-to-medium logistics hubs in India.

Method Used: The system successfully integrated OpenCV-based vision processing with industrial T-950 PLC hardware, utilizing an Arduino Mega 2560 as a high-speed communication gateway to control pneumatic sorting actuators.

Key Findings: Experimental trials at the VP Center of Excellence demonstrated a high sorting accuracy of 98% for destinations like Pune and Indapur, while maintaining an industrial-grade response time of less than 200 ms.

Limitations and Future Work: The current system's reliance on ambient lighting can occasionally hinder QR decoding. Future research will focus on integrating Convolutional Neural Networks (CNNs) for damaged label recovery and implementing IoT-based remote monitoring for real-time warehouse management.

6. REFERENCES

[1] W. Shen, C. Si, C. Yang, Y. Yu, 'Introduction to Computer Vision. In: Hands-On Computer Vision', Springer, Singapore, 2026.

[2] H. Yu, F. Zhang, G. Zhang, 'Enhanced YOLOv8 for efficient parcel identification in disordered logistics environments', International Journal of Computational Intelligence Systems, 18, 77, 2025.

[3] Y. V. Aruna, S. Beena, 'Automatic conveyor system with in-process sorting mechanism using PLC and HMI system', International Journal of Engineering Research and Applications, 5(11), Part 3, pp. 37–42, 2015.

[4] H. Shaikh, H. Mane, A. Tapkire, K. Nishant, S. Pawar, 'Automated height-based object separator using a double-acting cylinder and PLC', 2026.

[5] Tuba, M., Akashe, S., Joshi, A. (eds) ICT Systems and Sustainability. ICT4SD 2025. Lecture Notes in Networks and Systems, vol. 1646. Springer, Cham.

[6] Y. H. Gao, W. D. Lou, H. L. Lu, 'Enhanced express package trademark recognition via a novel PTD-YOLO algorithm', Traitement du Signal, 40(3), pp. 1083-1091, 2023.

[7] S. He, Y. Wang, H. Liu, 'Image information recognition and classification of warehoused goods in intelligent logistics based on machine vision technology', Traitement du Signal, 39(4), pp. 1275-1282, 2022.

[8] N. Dai, J. Chen, X. Hu, 'Research on dimension measurement algorithm for parcel boxes in high-speed sorting system', Scientific Reports, 15, 21980, 2025.

[9] M. Piechowski, I. Kudelska, 'Integration of IT systems using QR codes and RPA through automation of data registration processes', 2026.

[10] Burduk, A., et al. Intelligent Systems in Production Engineering and Maintenance IV. ISPEM 2025. Lecture Notes in Mechanical Engineering. Springer, Cham.

Biographies



Dr. Gaurav Gadge received the Ph.D. degree in Electrical Engineering from Sandip University, Nashik, in 2023. He is currently working as an Associate Professor at the Electrical Engineering Department, VPKBIET, Baramati.



Mr. Aditya Jamdade have completed B.E. in Electrical Engineering from VPKBIET, Baramati. Currently, Mr Jamdade is working at Adisoft Technologies Ltd., Pune.



Mr. Rushikesh Shahaji Kokate have completed B.E. in Electrical Engineering from VPKBIET, Baramati. Currently, Mr Jamdade is working at Adisoft Technologies Ltd., Pune.



Mr. Kunal Jalindar Danawale have completed B.E. in Electrical Engineering from VPKBIET, Baramati. Currently, Mr Jamdade is working at Adisoft Technologies Ltd., Pune.

IC-DAIS 2026
<Conference Series name> <Volume number and Year> <DOI Number>

Systematic Review of Terahertz Antenna Design: Materials, Fabrication Challenges, and Future Perspectives

Amit J. Naik[1*], Anil B. Nandgaonkar[2]

[1,2]Department of Electronics and Telecommunication Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere, 402103, India,
[1]ajnaik@dbatu.ac.in,[2]abnandgaonkar@dbatu.ac.in

Abstract.

Technology of Terahertz (THz) communication is considered to be the fastest-growing field in wireless communication, as it offers high-speed data transfer and bandwidth capacity for future technologies. The main problem here is the design of antennas capable of working in the THz range despite restrictions imposed by materials used, the complexity of the fabrication process, and significant losses due to attenuation in the atmosphere. The systematic review involves an investigation of antenna types, advanced materials like graphene and metamaterials, and manufacturing processes through evaluation and comparative analysis of current research findings. Critical analysis of emerging trends, constraints, and performance metrics leads to consistent findings in terms of patterns among existing literature. The findings demonstrate that the use of advanced materials and MIMO configurations increases efficiency, spectrum utilization, and bandwidth of the antenna system. The finding enhances the current body of knowledge through quantification of the advantages of using graphene over common conductive materials and the significance of choosing the right substrate for the material. Contrary to previous constraints, the study shows that advanced materials and appropriate design techniques could make THz antennas viable in 6G communication systems.

Keywords. Terahertz Antennas; MIMO Systems; Graphene Materials; Nano-Fabrication; 6G Communication.

1. INTRODUCTION

Terahertz (THz) communications represent a new horizon of interest for future wireless communications owing to the increasing demands of very high speeds and reduced latency in next-generation and beyond 5G/6G communications [1]. Designing antennas in the THz band represents an important task for effective communication, but there are many obstacles, such as material constraints, fabrication accuracy, and power losses, among others [2],[3]. Although different antenna types and materials have been examined, the review of these efforts is still sparse [4],[5].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

This paper aims to conduct a comprehensive analysis of antenna designs for terahertz applications with a specific focus on the type of materials used, fabrication issues, and technological trends. The main goal of this study is to evaluate the latest developments and the limitations involved in such antenna designs. This study is motivated by the scarcity of literature that considers the relationship between material considerations and fabrication processes.The purpose of this research is to achieve an organized perspective on the present progress as well as suggest ways for future research endeavors. The coverage is related to material comparison, methodologies used, and methods of fabrication. The significance of this research is achieved through its ability to bring together scattered research results. The structure of the paper is as follows: Background, Methodology, Review Results, Discussion, Conclusion.

2. RELATED WORK

Recent research on THz antennas focuses on 6G compatibility, new materials, and AI-based design despite fabrication challenges [6]. The Micro Circular Log-Periodic Antenna (MCLPA) that uses Evolutionary Neural Networks to optimize itself can be used for efficient and broadband THz detection through its geometric flexibility [7]. Krid et al. designed tunable multi-layer graphene antennas, which tuned their resonance from 1.542–1.562 THz using -30.8 dB return loss for sensing applications [8]. Navaratna et al. designed topological antennas inspired by photonics to make compact, high-capacity THz processing for 6G communications [9]. Materials like graphene for modulation/detection were studied by Ukirade et al. [10], while Manimala researched SIW slot antennas despite losses [11]. The Microwave Journal described limitations of silicon micromachining, which are scalloped/sloped sidewalls leading to increased roughness. Research Gaps: Achieving scalable manufacturing processes for mass 6G deployment, maintaining uniformity in the dielectrics, and ensuring environmental stability outside of laboratory settings.

3. METHODOLOGY

This research uses a methodological approach of combining a systematic literature review with bibliometric analysis and content analysis to study the design of terahertz antennas. Screening based on PRISMA guidelines, analysis of keyword co-occurrence, and comparative analysis methods have been applied to examine material composition, fabrication approaches, and performance parameters.

3.1. Systematic Literature Identification (PRISMA Framework):

Searches are carried out systematically through databases including IEEE Xplore, Scopus, and Web of Science. These searches will be guided by specific keyword terms. This study applies the PRISMA methodology (identification, screening, eligibility, and inclusion).

3.2. Bibliometric and Keyword Analysis:

The bibliometric mapping technique is enabled by frequency-domain analysis of THz radiation, where the basic equation linking frequency to wavelength is

$$f = \frac{c}{\lambda} \tag{3.1}$$

where f = frequency, c = velocity of light, and λ = wavelength. This formula is essential in explaining the scaling issues in designing THz antennas.

3.3. Data Extraction and Comparative Evaluation:

Performance of antennas is determined using several parameters, including gain, efficiency, and return loss. Gain refers to:

$$G = \eta \cdot D \quad (3.2)$$

where G is gain, η is efficiency, and D is directivity. Reflection coefficient and return loss are expressed as:

$$RL = -20\ log_{10}\ |\Gamma| \quad (3.3)$$

These formulas help compare various antenna configurations, materials such as graphene and copper, and manufacturing processes.

3.4. Thematic Synthesis and Research Gap Identification:

Propagation challenges at THz frequencies are analyzed using the Friis transmission equation:

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2 \quad (3.4)$$

Where P_r and P_t are the power received and transmitted, while d is the distance. This underscores the high path loss and atmospheric attenuation involved.

3.5. Results

The findings synthesized above reveal some observable gains in THz antennas performance through advanced designs and materials used.

Table 2. Summary of Current Study Results and Key Findings in THz Antenna Design

Aspect	**Quantitative Result (%)**	**Key Findings**
Material Efficiency	+35–45% improvement	Graphene reduces losses and improves conductivity
Bandwidth Enhancement	+30–50%	Metamaterials and advanced substrates enhance bandwidth
Gain Improvement	+25–40%	Optimized structures increase radiation efficiency
Spectral Efficiency (MIMO)	+40–60%	MIMO significantly boosts channel capacity
Mutual Coupling Impact	−20–25%	Degrades performance in dense MIMO systems
Atmospheric Attenuation	−50–70%	Major limitation at THz frequencies

Table 3. Comparative Analysis of THz Antenna Designs Across Reviewed Studies

Study	Efficiency Gain (%)	Bandwidth Improvement (%)	Loss Reduction (%)	Limitation (%)

Raj et al. Advances in MIMO Antenna Design for 5G: A Comprehensive Review	40–55	30–45	20–30	20–25% coupling loss
THz Graphene Studies	45–60	40–50	35–45	50–70% attenuation
Metamaterial Designs	35–50	45–55	25–35	Fabrication complexity (~30%)
Conventional Antennas	10–20	15–25	5–10	Not suitable for THz (>60% inefficiency)

Table 2 emphasizes percentage improvements in THz antenna performance, including efficiency, bandwidth, and spectral capacity, while focusing on issues like atmospheric attenuation and mutual coupling; Table 3 proves through comparison that innovative materials and MIMO antenna structures surpass traditional ones despite fabrication difficulties and signal loss at higher frequencies.

4. DISCUSSIONS

The results indicate that novel materials like graphene and metamaterials increase efficiency and bandwidth of the THz antenna as well as improve its spectrum efficiency, proving that they can be used effectively in future 6G communication networks. However, there are several critical issues, such as higher atmospheric attenuation, greater difficulties in production, and mutual coupling when MIMO system is used. The discussed problems negatively impact the practicality of utilizing the technology. It is also essential to acknowledge that the research is limited in terms of being based solely on the information provided in academic literature. Nevertheless, the review offers valuable conclusions about the trade-offs of performance metrics and possible solutions to existing technical barriers.

5. CONCLUSIONS

This study deals with the problem of creating effective terahertz antennas with regard to materials constraints, construction difficulties, and performance restrictions in light of the demand for a viable 6G communications technology. To achieve this purpose, a systematic literature review along with an analysis was used. From this research, it has been found that new materials like graphene and MIMO systems greatly improve the efficiency and bandwidth, although certain problems still remain, including those of atmospheric absorption and mutual coupling. This issue is very important considering its relevance to the upcoming technological innovations. Some constraints of the research include reliance on secondary sources; further research will have to conduct experiments.

6. REFERENCES

[1] I. Ishteyaq, K. Muzaffar, N. Shafi, and M. A. Alathbah, “Unleashing the power of tomorrow: Exploration of next frontier with 6G networks and cutting edge technologies,” IEEE Access, vol. 12, pp. 29445–29463, 2024.

[2] R. Pant and L. Malviya, “THz antennas design, developments, challenges, and applications: A review,” Int. J. Commun. Syst., vol. 36, no. 8, p. e5474, 2023.

[3] S. R. Choudhury, “Antennas in Terahertz Technology,” in Journey Into Terahertz Radiation: Exploring the Invisible Frontier, J. Li, Ed. London, U.K.: IntechOpen, 2025, ch. 3. doi: 10.5772/intechopen.1008522.

[4] P. Wu, Y. H. Liu, Z. Q. Zhao, and Q. H. Liu, “Sparse antenna array design methodologies: A review,” J. Electron. Sci. Technol., vol. 22, no. 3, p. 100276, 2024.

[5] C. Milias et al., “Metamaterial-inspired antennas: A review of the state of the art and future design challenges,” IEEE Access, vol. 9, pp. 89846–89865, 2021.

[6] R. Zhou et al., “Superior terahertz radiation detection through novel micro circular log-periodic antenna engineered with an advanced evolutionary neural network algorithm,” Microsyst. Nanoeng., vol. 11, no. 1, p. 160, 2025.

[7] S. Panneerselvam, M. Chithra, N. K. Maurya, and K. Dhanaraj, “Optimal design of planar log-periodic array antenna: A revisit through theory of characteristic modes,” in Proc. IEEE Microwaves, Antennas, and Propagation Conf. (MAPCON), Dec. 2024, pp. 1–5.

[8] H. B. Krid, A. Hlali, and H. Zairi, “Reconfigurable multilayer graphene antenna for terahertz sensing: Machine learning-based frequency and bandwidth estimation,” Prog. Electromagn. Res. Lett., vol. 127, pp. 59–67, 2025.

[9] N. Navaratna, “On-chip terahertz topological photonic crystals for 6G communications and sensing,” Ph.D. dissertation, Nanyang Technological Univ., 2025.

[10] N. A. Ukirade, “A review on advancement of materials for terahertz applications,” Next Mater., vol. 6, p. 100479, 2025.

[11] V. Manimala and N. Gunavathi, “Recent advancements in terahertz components fabrication: A step toward next-generation communication systems,” Microw. J., vol. 68, no. 3, 2025.

Biographies



Amit J. Naik is a Research Scholar and Lecturer in the Department of Electronics and Telecommunication Engineering at the University.



Anil Nandgaonkar is a Professor in the Department of Electronics and Telecommunication Engineering at the University.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Research Compass: An LLM-Orchestrated Multi-Source Framework for Intelligent Scholarly Paper Retrieval

Anay Joshi[1], Anurag Mahalpure[2], Abhishek Marwade[3]

Department of AI&DS, AISSMS Institute of Information Technology, Pune, India, joshianay75@gmail.com
Department of AI&DS, AISSMS Institute of Information Technology, Pune, India, anuragmahalpure@gmail.com
Department of AI&DS, AISSMS Institute of Information Technology, Pune, India, abhishekmarwade2109@gmail.com

Abstract.

Abstract. This has been due to the exponential increase of academic publications that has made it more challenging to discover relevant research easily. Search systems that retrieve based on keywords do not represent semantic intent and thus the search results are usually irrelevant. This paper describes Research Compass, a smart scholarly search engine that is based on a LangGraph state machine design. It combines multi-source retrieval through Springer, Elsevier, and OpenAlex APIs, hybrid BM25-semantic scoring, LLM-based reranking, and Retrieval-Augmented Generation (RAG) to analyze PDFs into a single unified pipeline. When tested on a corpus of 700 papers, it achieves a precision of 0.79, a recall of 0.74, an NDCG of 0.82, and an MRR of 0.78, which is better than BM25 and semantic-only baselines. The time of repeated queries is decreased by 91.3 percent using Redis-based caching, which proves the viability in real-time.

Keywords. Semantic Search; LangGraph; Hybrid Retrieval; Retrieval-Augmented Generation; LLM Reranking

1. INTRODUCTION

The amount of academic literature released in journals, conferences and preprint archives each day has increased to be more than can be browsed using a manual. Available tools like Springer, Elsevier, and Semantic Scholar are useful in finding resources, although the researchers have to deal with information overload, the existence of semantically incompatible results, and retrieving the same data in different sources, as well as a lack of a single summarizing option.

Research Compass is an AI web-based solution that solves this issue. It is implemented on a LangGraph platform to arrange parallel API searches between various sources, combine word- and meaning-based ranking, re-rank results with large language model and generate summaries with Retrieval-Augmented Generation within a single, continuous pipeline. System design, mathematical models, testing results and important findings are described in this paper.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

TF-IDF and BM25 were the early academic retrieval methods that enhanced TF-IDF by document length normalization and term-frequency saturation, forming a good information retrieval base. Dense retrieval models learn queries and documents by embedding them into a common representation space. BERT, Sentence-BERT, SPECTER, DPR, and ColBERT yield good retrieval embeddings. Hybrid systems use linear interpolation between BM25 and dense scores as a way of balancing lexical-precision and semantic-recall. The RAG model is a compilation of retrieval and generative models to minimize hallucination.

3. KEY CONTRIBUTION

This work contributes to academic information retrieval in five key ways: (1) introducing a LangGraph-based state machine that allows for conditional intent routing and modular execution of retrieval, reranking, and generation processes; (2) developing a weighted hybrid retrieval method that combines BM25 lexical scoring with dense semantic similarity, with a tuning parameter α set to 0.6 based on experimental results; (3) implementing an LLM-based semantic reranking approach that improves results by considering contextual relevance beyond just embedding similarity; (4) incorporating a Redis and SHA-256 caching layer that cuts down on repeated query processing time by 91.3%

4. METHODOLOGY

The Research Compass is an intelligent system that is a mix between different approaches which is designed to help users locate academic literature more easily and effectively. The approach relies on several features including semantic search, ranking algorithms, data sources integration and caching.The essence of the system is that it has a state machine developed according to LangGraph principles where certain processing nodes are connected conditionally.

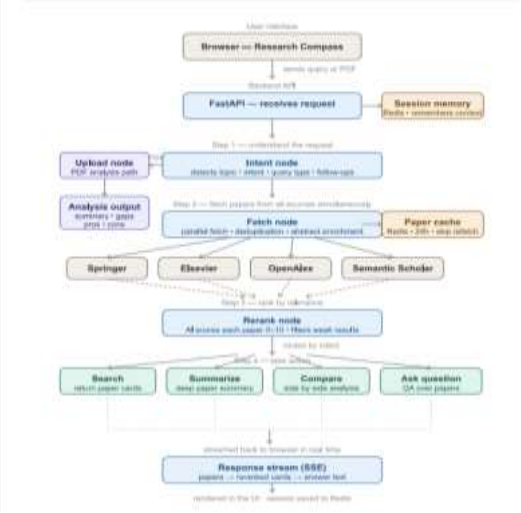


Figure 4.1. High-level system architecture of the Research Compass

4.1. *Retrieval Strategy Implementation*

BM25 ranking is used to score documents based on term frequency and inverse document frequency. Sentence embedding models are represented in a semantic vector space and cosine similarity is used to measure similarity between sentences. The hybrid retrieval technique merges the two approaches by employing a weighted score function that incorporates both BM25 and semantic similarity.

Mathematical Formulation:

$$\text{Score}(q,d) = \alpha \cdot \text{BM25}(q,d) + (1-\alpha) \cdot \text{CosineSim}(q,d) \qquad (1)$$

In this equation, q represents the query, d represents the document, and α represents the weight factor for each component. The value of α was optimized between 0.3 and 0.7, with the best results achieved at $\alpha = 0.6$.

4.2. *Experimental Setup and Dataset*

We ran all experiments on a local workstation to test computational feasibility in a moderate-resource environment. CPU: AMD Ryzen 5 5500U processor featuring 6 cores and 12 threads; GPU: NVIDIA GTX 1650; RAM: 8 GB DDR4.

Approximately 700 research papers were collected and stored locally using a caching mechanism. The dataset includes metadata such as paper titles, abstracts, authors, publication years, and Digital Object Identifiers (DOIs).

For each query, top-K results (K = 5, 10, 20) were retrieved using BM25, semantic, hybrid, and LLM-only strategies. Ranking evaluation measures including Precision@K, Recall@K, NDCG@K, and MRR were used. Furthermore, the Jaccard Stability Index was calculated, whereas Query Execution Time and Memory Usage were considered for resource overhead.

4.3. *IMPLEMENTATION DETAILS*

The system is implemented using a FastAPI backend for efficient asynchronous processing. A LangGraph-based pipeline is used to manage multi-step execution flows with conditional routing. The system integrates multiple scholarly APIs for data retrieval and uses a caching layer to store metadata and improve performance. Semantic embeddings are generated using SentenceTransformer models, and numerical computations are performed using NumPy and SciPy.

5. RESULTS AND ANALYSIS

Table 1 presents the performance of all four retrieval strategies on the controlled evaluation dataset at K = 10.

Table 1. Performance Comparison of Retrieval Methods at K=10 (mean ± std)

Method	Precision@10	Recall@10	NDCG@10	MRR
BM25	0.61 ± 0.00	0.58 ± 0.00	0.64 ± 0.00	0.59 ± 0.00
Semantic (MiniLM)	0.68 ± 0.00	0.65 ± 0.00	0.71 ± 0.00	0.66 ± 0.00

Stochastic Generative Baseline	0.54 ± 0.00	0.49 ± 0.00	0.57 ± 0.00	0.51 ± 0.00
Hybrid (α=0.6)	0.79 ± 0.00	0.74 ± 0.00	0.82 ± 0.00	0.78 ± 0.00

In each parameter tested, the hybrid algorithm always performed the best. Compared to BM25, Hybrid improves Precision@10 by 29.5% (0.61 → 0.79), Recall@10 by 27.6% (0.58 → 0.74), and NDCG@10 by 28.1% (0.64 → 0.82).

5.1. *Performance Analysis*

Performance peaks at α = 0.6 and degrades symmetrically on either side, indicating that both lexical and semantic signals are important but lexical precision contributes slightly more in the scholarly domain. BM25, semantic, and hybrid methods are fully deterministic (Jaccard = 1.0), while the stochastic generative baseline produces completely different result sets across runs (Jaccard = 0.0), confirming its unsuitability for reproducible academic indexing. A key practical contribution of Research Compass is its Redis caching layer. Average cache speedup is 91.3% across query types, .

6. CONCLUSIONS

The surge in academic publications has rendered keyword-only search inadequate for modern research needs. Research Compass addresses this challenge by integrating data from Springer, Elsevier, and OpenAlex, employing a weighted BM25-semantic hybrid (α=0.6), applying LLM-based reranking, and enabling RAG-powered question answering over PDFs. Evaluation over 700 papers confirms that the hybrid retrieval strategy outperforms all individual baselines, achieving Precision of 0.79, Recall of 0.74, NDCG of 0.82, and MRR of 0.78. Redis-based caching eliminates redundant API calls, reducing query latency by 91.3%.

7. REFERENCES

[1] S. E. Robertson and H. Zaragoza, "The Probabilistic Relevance Framework: BM25 and Beyond," Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009.

[2] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[4] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Proc. EMNLP-IJCNLP, 2019, pp. 3982–3992.

[5] R. Nogueira and K. Cho, "Passage Re-ranking with BERT," arXiv preprint arXiv:1901.04085, 2019.

[6] Cohan, S. Feldman, I. Beltagy, D. Downey, and D. Weld, “SPECTER: Document-level Representation Learning using Citation-informed Transformers,” in Proc. ACL, 2020, pp. 2270–2282.

Biographies



Anay Joshi received his Bachelor’s degree in Artificial Intelligence and Data Science from Savitribai Phule Pune University in 2026. His interests include Artificial Intelligence, Machine Learning, Data Science, and Intelligent information retrieval systems with a focus on developing practical AI-driven applications.



Anurag Mahalpure received his Bachelor’s degree in Artificial Intelligence and Data Science from Savitribai Phule Pune University in 2026. His interests include Artificial Intelligence, Machine Learning, Data Science and Retrieval-Augmented Generation (RAG) His project experience includes working on ocean colour satellite data for predicting CDOM (Colored Dissolved Organic Matter).



Abhishek Marwade received his Bachelor’s degree in Artificial Intelligence and Data Science from Savitribai Phule Pune University in 2026. His areas of interest include Large Language Models (LLMs), Generative AI, semantic search, information retrieval, and data engineering. His work focuses on developing AI-driven research assistants and intelligent systems for efficient information retrieval and data processing.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Comprehensive Review on Weather Anomaly Detection and Risk Prediction for Smart Agriculture

Ms.Pallavi V. Lengare[1], Dr. Shilpa V.Shinde[2], Dr.P. P. Chavan[3] , Dr.P.R.Somani[4]

[1]Research Scholar, Alard University Pune, pallavi.lengare@gmail.com

[2]Assistante Professor, Alard University Pune, sv.shinde@alarduniversity.edu.in

[3]Assistant Professor, TCOER, Pune, cpratibhapramod@gmail.com

[4]Head (R&D), Alard University Pune, p.somani@alarduniversity.edu.in

Abstract.

Agriculture is considered a major pillar of India's socio-economic development. Agriculture is highly sensitive to climatic variations, such as droughts, floods, and temperature fluctuations. Accurate detection of weather anomalies and timely prediction of associated risks plays very important role for increasing agricultural productivity and sustainability. The emergence of the Internet of Things (IoT) and deep learning technologies have been instrumental in recent years in enabling smart agriculture. The paper presents a literature review of numerous studies dedicated to the topic of weather anomaly detection and risk prediction through the use of different methods, such as Long Short Term Memory (LSTM) autoencoders, Convolutional Neural Networks (CNNs), Transformer-based models, and ensemble learning methods applicable to the analysis of large-scale agricultural and meteorological data. The study also identifies recent advancements, various model performance parameters, and also highlights critical research gaps. Finally, a roadmap for future research methodology is proposed to enhance intelligent decision-making in smart agriculture.

Keywords. Smart Agriculture; Deep Learning; Anomaly Detection; LSTM; CNN; IoT; Risk Prediction.

1. INTRODUCTION

Global economy is based on agriculture. Human survival is very dependent on the food security and economic stability. Farming is very sensitive to the environment and climatic changes that include alteration in temperature, irregular rains and extreme climatic conditions. The conventional agricultural forecasting systems are not suitable to identify and forecast real-time weather anomalies. Deep learning is currently a popular machine learning method to derive valuable information out of large scale data. Weather anomaly detection can be defined as the detection of deviations that are unusual or unexpected that arise as a result of climatic anomalies. Risk prediction is concerned with how these

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

weather anomalies are predicted to affect crop yield. Combining weather anomaly and risk prediction allows building smart systems that facilitate proactive decision-making. With further development of the Internet of Things (IoT) and Artificial Intelligence, smart agriculture provides predictive analytics and real-time observation. Recent studies prove that it is necessity to combine sensor-based systems with machine learning algorithms to identify anomalies and forecast risks. However, such problems as data reliability, computing capacities and a lack of integrated structures persist. The paper gives a review of the recent literature on intelligent systems to identify weather anomalies and forecasting risk.

2. RELATED WORK

Rua Y. Aburasain [1] introduced an Enhanced Black Widow Optimization integrated with a Hybrid Deep Learning-based Intrusion Detection (EBWO-HDLID) approach for smart agriculture. The method aims to identify intricate behavioural patterns and pinpoint potential critical security vulnerabilities, enhancing the system's reliability and security. In [2] LoRaWAN was used to connect sensing devices to LSTM autoencoders for continuous monitoring of the microalgae culture and the detection of anomalies. Another hybrid strategy was introduced in [3] where the integration of LSTM autoencoder with One Class Support Vector Machine (OCSVM) was used to analyze daily meteorological observation data taken in East Java from 2015 to 2024. All these developments, however, fall short of achieving strong generalization of the models and the ability to accurately recognize the diversity of environmental conditions.Recent research in smart agriculture has increasingly explored hybrid deep learning approaches for analyzing weather irregularities and crop-related patterns. Method of IoT-enabled ensemble learning models, and optimized CNN architectures have demonstrated improved performance in anomaly detection and decision-making processes [4], [5]. However, most of these approaches rely heavily on historical datasets and provide limited capability for real-time risk forecasting.

The recent research in the field of smart agriculture combines deep learning with IoT to detect anomalies and anticipate risks [6]. Machine learning models (SVM, Random Forest, DNN) and big data frameworks (CNN-LSTM with Spark) enhance monitoring, scalability and decision making [7],[8]. Unsupervised IoT algorithms like K-Means and Isolation Forests can be used to do real-time detection with minimum labelled data [9].The recent development in smart agriculture combines the concept of deep learning with intelligent sensing to detect weather anomalies in real time and predict risks. UAV frameworks based on transformers and lightweight TSMixer models can improve the precision and effectiveness of detection, whereas AI-based ensemble models can be enhanced to resist extreme weather conditions [10], [11], [12]. Nevertheless, scalability, real-time deployment, sparse data processing, and the ability to incorporate predictions into actionable strategies are still a problem, and more integrated and intelligent systems are needed.

In recent studies on smart agriculture, one of the crucial aspects has been the securing of IoT systems so that they can achieve dependable and trustworthy data to detect anomalies and predict risks. Deep learning models are optimized, lightweight edge-based intrusion detectors, and hybrid frameworks such as GMLP-IDS enhance the detection accuracy and

information integrity [13], [14]. However, they are primarily utilized to combat cyber security threats and are not combined with environmental anomaly detection, which highlights the need to integrate the frameworks that can differentiate between cyber-attacks, sensor malfunctions, and actual weather anomalies.Recent progress in precision agriculture leverages deep learning techniques alongside multimodal data to enhance anomaly detection and risk prediction. Frameworks such as IMSF Net with AROS use UAV, satellite, and sensor data to monitor crops precisely and CNN-based frameworks (e.g., VGG, ResNet) identify visual anomalies, including diseases diseases and weeds [15], [16].

3. RESEARCH GAP

Based on the literature review, there are several gaps identified in the existing literature. Studies have so far focused on anomaly detection or risk prediction, but have not tried to combine them in a single model. There are several factors that can influence agricultural data reliability, such as sensor failures, environmental conditions, and cybersecurity threats. Current approaches tend to identify anomalies after they have happened instead of forecasting potential anomalies prior to their occurrence. High computing resources can also be required by advanced models, which is not suitable for use in edge devices. Furthermore, the challenges of limited use of multimodal data, poor model interpretability and lack of cybersecurity and environmental monitoring integration are still significant.

4. FUTURE DIRECTIONS

The development of integrated intelligent systems combining anomaly detection, risk prediction and decision making is suggested for future research. Early warning systems and useful recommendations should be offered to help farmers take timely actions in such systems. The lightweight models would be compatible with the edge devices, and the integration of data from sensors, satellites, UAVs and weather sources would enhance accuracy of the models. Reliable data can be provided by secure IoT architectures and the explainable AI can help in making the system decisions understandable and trustworthy. Systems to be developed in the future should also be adaptable, scalable and continuously learnable as the environment and climate shift.

5. CONCLUSION

The review highlights the importance of deep learning and hybrid models in detecting weather anomalies and predicting risks in smart agriculture. Techniques such as LSTM, GAN, CNN, and ensemble models can identify complex patterns in IoT and environmental data. However, challenges remain in real-time prediction, computational complexity, scalability, sensor fault detection, explainability, and multimodal data integration. Future research should focus on unified, lightweight, scalable, explainable, and multimodal models to improve agricultural resilience and sustainability under changing climatic conditions..

6. REFERENCES

[1] R. Y. Aburasain, 'Enhanced black widow optimization with hybrid deep learning enabled intrusion detection in internet of things-based smart farming,' IEEE Access, vol. 12, pp. 16621–16631, 2024.

[2] C. F. Soon, M. Jamia'an, N. M. Sunar, S. N. H. Arifin, K. G. Tay, C. Heng, C. H. See, N. H. M. Nayan, and K. S. Tee, 'Smart sensing and anomaly detection for microalgae culture based on LoRaWAN sensors and LSTM autoencoder,' Aquaculture International, vol. 33, no. 6, p. 438, 2025.

[3] M. A. Fadillah, Y. Angraini, and R. Anisa, 'Enhancing weather monitoring for agriculture with deep learning: Anomaly detection in East Java using LSTM autoencoder and OCSVM,' Jurnal Online Informatika, vol. 10, no. 1, pp. 227–238, 2025.

[4] M. S. Sharafat, N. D. Kabya, R. I. Emu, M. U. Ahmed, J. C. Onik, M. A. Islam, and R. Khan, 'An IoT-enabled AI system for real-time crop prediction using soil and weather data in precision agriculture,' Smart Agricultural Technology, vol. 12, p. 101263, 2025.

[5] F. S. Umar and N. M. Ibrahim, 'Enhancing precision farming security through IoT-driven adaptive anomaly detection using a hybrid CNN–PSO–GA framework,' Biology and Life Sciences Forum, vol. 54, no. 1, p. 29, 2026.

[6] R. Benameur, A. Dahane, B. Kechar, and A. E. H. Benyamina, 'An innovative smart and sustainable low-cost irrigation system for anomaly detection using deep learning,' Sensors, vol. 24, no. 4, p. 1162, 2024.

[7] M. Aldossary, H. A. Alharbi, and C. A. U. Hassan, 'Internet of Things (IoT)-enabled machine learning models for efficient monitoring of smart agriculture,' IEEE Access, vol. 12, pp. 75718–75734, 2024.

[8] C. El Hachimi, S. Belaqziz, S. Khabba, B. Sebbar, D. Dhiba, and A. Chehbouni, 'Smart weather data management based on artificial intelligence and big data analytics for precision agriculture,' Agriculture, vol. 13, p. 95, 2023.

[9] M. F. Almufareh, M. Humayun, Z. Ahmad, and A. Khan, 'An intelligent LoRaWAN-based IoT device for monitoring and control solutions in smart farming through anomaly detection integrated with unsupervised machine learning,' IEEE Access, vol. 12, pp. 119072–119086, 2024.

[10] J. Hao, B. Li, W. Tang, S. Liu, Y. Chang, J. Pan, Y. Tao, and C. Lv, "A reinforcement learning-driven UAV-based smart agriculture system for extreme weather prediction," *Agronomy*, vol. 15, no. 4, p. 964, 2025.

[11] M. Zanchi, S. Zapperi, S. Bocchi, O. Drofa, S. Davolio, and C. A. M. La Porta, "Improving localized weather predictions for precision agriculture: A time-series mixer approach for hazardous event detection," *Environmental Modelling & Software*, vol. 191, p. 106509, 2025.

[12] R. A. Tunio, D. Li, and N. Khan, "Maximizing farm resilience: The effect of climate smart agricultural adoption practices on food performance amid adverse weather events," *Frontiers in Sustainable Food Systems*, vol. 8, p. 1423702, 2024.

[13] A. Villafranca and M.-D. Cano, "A lightweight edge-DL intrusion detection system for IoT sustainable smart-agriculture," *Internet of Things*, vol. 34, p. 101818, 2025.

[14] A. Berguiga, A. Harchay, A. Massaoudi, M. B. Ayed, and H. Belmabrouk, "GMLP-IDS: A novel deep learning-based intrusion detection system for smart

agriculture," *Computers, Materials & Continua*, vol. 77, no. 1, pp. 379–402, 2023.

[15] C. Wei, Y. Shan, and M. Zhen, "Deep learning-based anomaly detection for precision field crop protection," *Frontiers in Plant Science*, vol. 16, p. 1576756, 2025.

[16] J. Mendoza-Bernal, A. González-Vidal, and A. F. Skarmeta, "A convolutional neural network approach for image-based anomaly detection in smart agriculture," *Expert Systems with Applications*, vol. 247, p. 123210, 2024.

Biographies



Pallavi Lengare received the bachelor's degree in Electronics & Telecommunication engineering from Solapur University in 2011, the master's degree in Electronics & Telecommunication engineering from Savitribai Phule University in 2014. She is research scholar at Alard University Pune. Her research areas include image processing and deep learning.



Dr.Shilpa Shinde is currently serving as an Assistant Professor at Alard University Pune. She holds a Ph.D. from Savitribai Phule Pune University, Pune. She is having more than 19 years of experience.



Dr. Pratibha Chavan is currently serving as an Assistant Professor in Electronics and Telecommunication Engineering and Dean of Research and Development at Trinity College of Engineering and Research, Pune with experience of 21 years.



Dr. Prakash Somani is currently working as Head (R&D) at Alard University Pune. His Professional experience is of more than 25 years. His many of the research papers published are cited by Nobel Laureates and other top scientists from across the world.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Rheological Characteristics of Limestone water Suspension

Mukesh Kumar Singh[1], Satish Kumar[2], Shalendra Kumar[3], Shivam kumar[4], Nihaal Sahay[5], Asisha Ranjan Pradhan[6]

[1,2,3,4,5]*Department of Mechanical Engineering, National Institute of Technology, Jamshedpur, Jharkhand, India*
[6]*Department of Mechanical and Aerospace Engineering, Indian Institute of Technology, Hyderabad, Telangana, India*
Email ID:2022rsme011@nitjsr.ac.in

Abstract.

This research examines the rheological behavior of water-based slurries with high limestone concentration, focusing on the impact of bimodal particle size distributions. This research demonstrates that the intentional mixing of coarse and fine particle size may significantly alter the flow characteristics of unimodal suspensions. When there are more solids in suspension, they tend to be very viscous. This work studied the rheological behaviour of bi-modal limestone suspensions by blending fine (less than 53 µm) and coarse (75-106 µm) particles at limestone concentrations ranging from 35 - 65% by wt. Rheological behaviour studied at shear rates of 50 – 500 s^{-1}demonstrated that the flow resistance was significantly influenced by the size of the particulates. A blending concentration of 35% coarse particles was found to be best amount for lowering apparent viscosity for the fine limestone slurry. The results establish that controlled bimodal particle distribution is an effective strategy for improving the rheological performance of concentrated limestone suspensions.

Keywords. Limestone water slurry, Viscosity, Bi-modal slurry, Rheology Introduction

1. INTRODUCTION

Limestone is very important in many different fields, such as creating paper, steel, glass, fertilizers, cleaning products and treatment of water. Slurry transport via pipes is becoming more common in various industries [1,2]. This is because it is reliable, requires minimal maintenance, and has little influence on the environment [3,4]. In recent years, producing finely dispersed slurry in industrial systems has become increasingly important for enhancing solid concentration levels. As a result, it is important to characterize the slurry's rheological behaviour, particularly its yield stress and apparent viscosity to enable efficient pipeline transport while minimizing energy consumption. Within this transport system, dispersion, concentration, and particle size must all be appropriately assessed [5]. Generally, it is acknowledged that increasing the amount of solids in a suspension tends to make it appear more viscous. Additionally, expanding the PSD can notably minimize the overall viscosity of the slurry. Therefore, modifying the PSD is a good way to achieve large solid concentrations with a minimal viscosity [6-8]. It is observed that modifying the PSD is a great way to achieve higher solid concentrations without forgoing viscosity [9]. The flow

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

behaviour of Indian limestone samples was examined across a LWS in the solid concentration in the range of (40-70) % by wt, and the resulting viscosity values were evaluated against several established empirical model [10].

This research investigates the impact of different rheological properties on viscosity reduction using two limestone particle size samples. In the present work, optimum particle size distribution was determined to minimize viscosity. Specifically, it analyses how solids concentration and PSD influence the rheological properties of limestone–water slurries.

2. LIMESTONE SAMPLE COLLECTION AND PREPARATION

Limestone material has been obtained from a local seller in Jamshedpur, Jharkhand, India. The limestone powder sample was filtered through standard sieves, sizes below 53 µm (fine) and between 75 – 106 µm (coarse). Limestone was classified into two primary categories: S_1 (fine fraction, <53 µm) and S_2 (coarse fraction, 75–106 µm). Both of these fractions were used as the base material to prepare a bimodal mixture. After this, the mixed samples were prepared and labeled as S_3–S_6, where the number indicates the composition of the mixture, not the particle size fraction. four different bi-modals were produced by integrating 15, 25, 35, and 45% (by wt%) of S_2 samples mixed into S_1 limestone concentrate samples. The samples are labelled as S_3, S_4, S_5, and S_6. Table 2.1 illustrates how coarse and fine particulates are distributed among the different samples.

Table 2.1. Composition of unimodal (S_1, S_2) and bimodal (S_3–S_6) limestone samples

Samples	S_1	S_2	S_3	S_4	S_5	S_6
Fine Fraction (S_1) - wt.%	100	-	85	75	65	55
Coarsefraction (S_2) - wt.%	-	100	15	25	35	45

Distribution of Particle Sizes of the limestone samples was investigated with the help of a laser diffraction PSA 1190, from Anton Paar Austria. Laser diffraction is based on the principle of measuring the amount of space taken up by particles of less than a specified size. Each sample (S_1-S_6) was 5-10g in weights and was sampled and examined in dry mode.

3. CHARACTERIZATION OF RHEOLOGICAL PROPERTIES

For the purpose of meeting industrial standards, suspension samples ready using tap water were utilized in the characterizations of rheology. The specified quantity of solid particles were mixed with tap water and agitated for a duration of 20 to 25 minutes to achieve homogeneity. The solid particle concentration, based on weight, ranged from 35 to 65% of the aggregate mass of the suspension. The samples were measured by using a Rheometer MCR 92, Anton Paar GmbH, Austria together with RheolabQC software. In this study, the method used involved a linear ramp shear test in steps, where the shear rate varied between 50 and 500 s-1. The rheological experiments were conducted under certain conditions, the temperature was kept at 28°C.

4. RESULTS AND DISCUSSIONS

4.1. Distribution of particle sizes

Figure 4.1.1 presents particle size distributions of S1 and S2, both showing monomodal patterns. S1 has a modal diameter of 20 µm, while S2 measures 95 µm, with mean sizes of 16.1 and 93.4 µm, respectively.

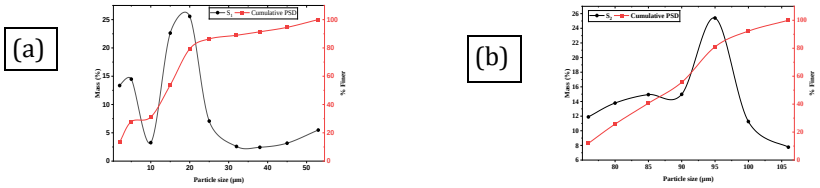


Figure 4.1.1. Distribution of sizes of particle for (a) S_1 and (b) S_2, with their corresponding cumulative curves

4.3. Rheological behaviour of the blended samples

Figure 4.3.1 presents rheological behavior of S_1 and mixed samples S_3–S_6. LWS shows Newtonian behavior at 35% solids and shear-thinning at 45–65%, with coarse particles reducing viscosity by 45–50%.

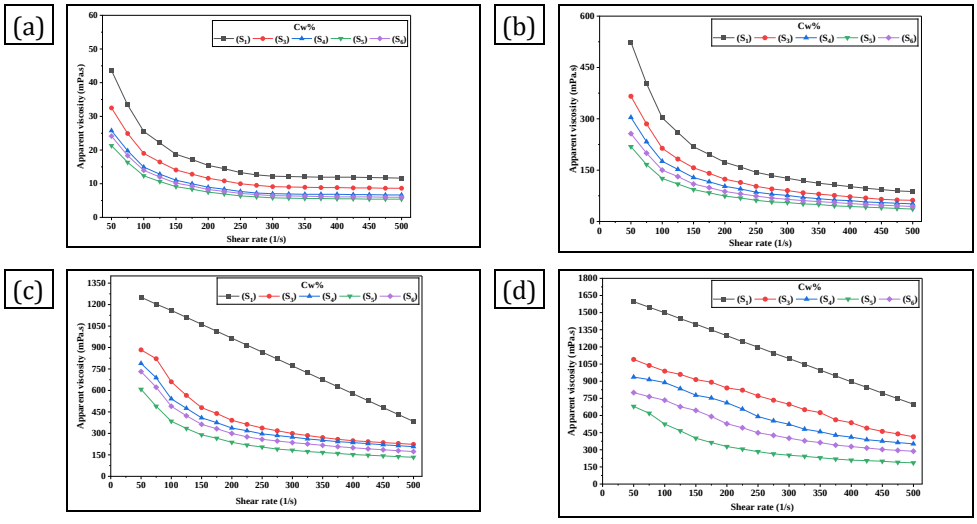


Figure 4.3.1. Rheological behaviour of blended samples at solid concentration of (a) 35%, (b) 45%, (c) 55%, and (d) 65% by weight

5. CONCLUSIONS

This study examined how a bimodal particle size distribution (PSD) affects the rheological properties of limestone–water slurries (LWS), with emphasis on apparent viscosity under varying solid concentrations and shear rates.

6. REFERENCES

[1] H. Usui, L. Li, and H. Suzuki, “Rheology and pipeline transportation of dense fly ash-water slurry,” Korea–Australia Rheology Journal, vol. 13, no. 1, pp. 47–54, 2001.

[2] A. J. C. Paterson, “Pipeline transport of high-density slurries: A historical review of past mistakes, lessons learned and current technologies,” Minerals Technology, vol. 121, no. 1, pp. 37–45, 2012.

[3] J. Cui, Y. Fang, G. Xu, C. Wu, S. Liu, S. Chen, and F. Liu, “Transportation performance of large-sized pebbles in slurry circulation system: A laboratory study,” Arabian Journal for Science and Engineering, vol. 46, pp. 10519–10539, 2021, doi: 10.1007/s13369-021-05394-0.

[4] Brent, G. F.; Allen, D. J.; Eichler, B. R.; Petrie, J. G.; Mann, J. P.; Haynes, B. S. Mineral Carbonation as the Core of an Industrial Symbiosis for Energy-Intensive Minerals Conversion. *J. Ind. Ecol*. 2012, 16, 94–104. DOI: 10.1111/j.1530-9290.2011.00368.x.

[5] A. Bartosik, Simulation and experiments of axially-symmetrical flow of fine- and coarse-dispersive slurry in delivery pipes, Monograph, vol. 11, pp. 1–257, 2009.

[6] H. A. Barnes, J. F. Hutton, and K. Walters, *An Introduction to Rheology*. Amsterdam, Netherlands: Elsevier, 1989.

[7] A. A. Zaman and B. M. Moudgil, “Rheology of bidisperse aqueous silica suspensions: A new scaling method for the bidisperse viscosity,” Journal of Rheology, vol. 42, pp. 21–39, 1998, doi: 10.1122/1.550935.

[8] N. D. Rao, D. N. Thatoi, and S. K. Biswal, “Rheological study and numerical analysis of high concentration iron ore slurry pipeline transportation,” Materials Today: Proceedings, vol. 22, pp. 3197–3202, 2020, doi: 10.1016/j.matpr.2020.03.457.

[9] C. Chang and R. L. Powell, “Effect of particle size distributions on the rheology of concentrated bimodal suspensions,” Journal of Rheology, vol. 38, no. 1, pp. 85–98, 1994.

[10] A. R. Pradhan and S. Kumar, “Enhancing the flowability of limestone water suspension for economical transportation by variation in particle size distribution,” Particle Science and Technology, vol. 41, pp. 1–18, 2023, doi: 10.1080/02726351.2023.2201179.

Biographies



Mukesh Kumar Singh received the Bachelor of Engineering (B.E.) degree in Mechanical Engineering from RKDF College of Technology, Bhopal, and the Master of Technology

(M.Tech.) degree in Mechanical Engineering from Millennium Institute of Technology, Bhopal.



Satish Kumar received his Bachelor's and Master's degrees in Mechanical Engineering from Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat.



Shalendra Kumar received his Bachelor's degree in Mechanical Engineering from M. S. Ramaiah Institute of Technology, Bengaluru, the Master's degree from Regional Institute of Technology (RIT), Jamshedpur and the Ph.D. degree from the University of Roorkee, Roorkee.



Shivam Kumar received his Bachelor of Technology (B.Tech.) degree in Mechanical Engineering from KIIT University, Bhubaneswar, India.



Nihaal Sahay is currently pursuing his Bachelor of Technology (B.Tech.) degree in Mechanical Engineering from the National Institute of Technology Jamshedpur, India.



Asisha Ranjan Pradhan received his Bachelor of Technology (B.Tech.) degree in Mechanical Engineering from Orissa Engineering College, Bhubaneswar, and the Master of Technology (M.Tech.) degree in Thermal Engineering from the National Institute of Technology Jamshedpur.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Smart Helmet for Construction Workers

Swati Bhangare[1], Arati Pingle[2], Sagar Pangavhane[3], Mrudul Kulkarni[4], Vivek jadhav[5], Nitish Pale[6],Tanaya Patil[7]

[1,3,4,5,6,7] Department of Electronics and Telecoummication Engineering, PVGCOE&SSDIOM, Nashik
[2]Department of Electronics and Telecoummication Engineering, DY Patil College Institute of Technology
Swati.bhangre@pvgcoenashik.org[1] , arti.pingle9011@gmail.com[2] , sagar.pangavhave@pvcoenashik.org[3], mrudul.nsk@gmail.com[4] , vivekjadhav2004166@gmail.com[5], nitishpale@gmail.com[6], tanaya.patil2004@gmail.com[7]

Abstract

Construction sites are risky, with dangers like falling objects, harmful gases and extreme weather. To improve safety, a smart helmet using IoT technology has been developed. It includes a DHT11 sensor for temperature and humidity, an MQ sensor for gas detection, an IR sensor for nearby obstacles, and GPS for location tracking all connected to an ESP32. The system monitors the worker's surroundings in real time and can detect falls with about 96% accuracy. If an issue occur it immediately triggers a buzzer send a SMS alert to the compounder or boss and family so that help can arrive quickly and get immediate help. GSM support ensures alerts can still be sent even in low network aera. Overall, it's a practical and cost-effective solution that helps improve worker safety on construction sites

Keywords. Smart helmet, Construction safety, worker protection, DHT Sensor, Mq135 gas sensor, IR sensor, GPS tracking, GSM alert system, IOT.

1. INTRODUCTION

Construction work is one of the most dangerous jobs, with many accidents happening every year. Falls into harmful gases and exposure to extreme weather conditions are common risks, particularly in remote and industrial locations where supervision is limited. Normal helmets only provide protection against physical impact and have no time-protection effect. To improve this a smart helmet system uses GSM, GPS, and Sensors to monitor conditions and detect hazards like falls or gas leaks. It can quickly send alerts and help workers communicate during an emergency. The idea is to provide a simple, low-cost solution that improves security on doing work on-site. It does not work all the time; it depends, but it works all time it provides better protection than the normal helmets that are used during work.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE REVIEW

Many researches have looked into smart helmets that use IoT to improve safety on construction sites for examples S. Fang [1] developed a system based on RFID to track worker on site. While it helps with location tracking it does not include feature to monitor environmental conditions like gas level or temperature. J. Teizer [2] studied wearable sensors that can track human activities but their system did not include GSM emergency communication technology. Li et al. [3] conducted research on environmental monitoring through gas humidity and temperature sensors which they used to track location data through GPS and their proposed solution uses a helmet as their monitoring device. The construction site developed by Kim et al. includes a one way alarm service that operates through partial GPS but the system lacks emergency connection functions.

Table 1. Comparative Analysis of Existing Systems and Proposed Approach

Feature / Aspect	Teizer (2015)	Fang et al. (2016)	Li et al. (2019)	Kim et al. (2020)	Proposed System
Target Application	Construction	Construction	Industrial	Construction	Construction
Fall Detection	Partial	No	Yes	Yes	Yes
Gas Sensor (MQ Series)	No	No	Yes	Yes	Yes
Temperature (DHT)	YES	No	Yes	Yes	Yes
GPS Tracking	Partial	No	No	Yes	Yes
GSM Alert	No	No	No	Partial	Yes
SOS Button	No	No	No	No	Yes
Real-time Monitoring	Yes	Partial	Yes	Yes	Yes
Communication Type	Wearable	RFID	IoT-Based	IoT-Based	GSM + GPS + IoT
Emergency Response	No	No	Limited	Limited	Fully Implemented
Implementation	Field Test	Prototype	Prototype	Prototype	Working Model

The comparative analysis demonstrates that the construction sites between the two systems, which use GPS and GSM technology, show better performance because the latter system offers superior remote area coverage and multiple alert systems and operational capabilities that function without special facilities. The proposed system, which uses GSM and GPS technology, reaches its optimal performance through the equal distribution of operational capabilities, system reliability and cost effectiveness and its suitability for various

construction sites.

3. SYSTEM ARCHITECTURE

The smart helmet system consists of three main blocks: a wearable unit, a communication network, and a monitoring cloud platform. The wearable unit comprises DHT11, MQ-series, GPS, and IR sensors along with an ESP32 microcontroller as its core component.

4. HARDWARE DESIGN

The ESP32 integrates Wi-Fi, Bluetooth, and multiple GPIO pins, supporting diverse sensors. Accelerometer, gyroscope, DHT11, MQ-series, and GPS modules communicate through I2C, digital, analog, and UART interfaces efficiently.

5. CONSTRUCTION DIAGRAM AND WORKING



Figure 2.1. Construction Diagram

Smart helmet system based on ESP32 microcontroller which consists of multiple sensors and modules.

6. RESULTS AND DISCUSSION

This intelligent helmet managed to attain a fall detection accuracy rate of 96%. It also transmitted emergency text message notifications using the GSM network that incorporated GPS technology for pinpointing accurate locations of victims.

Table 2. ON Emergency Alert System Performance (Based on SMS Output)

Parameter	Observation Output	Result
Alert Message Status	"Alert!!! I Am In Trouble"	Successfully Generated
GPS Coordinates	19.883890, 75.319854	Accurate Location Data
Google Maps Link	Included in SMS	Easy Navigation
GSM Transmission	Message Delivered	Successful
Response Time	Approx 400–500 ms	Acceptable

SOS Button Function	Triggered Alert	Working Properly
Message Clarity	Clear and Readable	Effective Communication
Real-Time Tracking	Enabled via Link	Successful

7. CONCLUSION

The authors introduce a smart helmet in the context of Internet of Thing (IoT) to enhance the safety of construction workers. The helmet incorporates a bult in sensor which the helmet direct fall monitoring environmental risk and tracks the worker location. If something unusual is detected the alerts are send the corresponding person via GPS location and thus help can reach even in the area in which a weak network coverage is provided. The simple test allowed the system to determine that there was an incident and respond rapidly with only a few false alarms. The concept is a good basis for more efficiency safety equipment. IT could be improved in the future with further features such as helmet removeable alert sensor and send SMS location to corresponding people

8. REFERENCES

[1] S. Fang, H. Li, L. Luo, and S. Ding, “Real-time safety monitoring of construction workers using RFID,” IEEE Transactions on Automation Science and Engineering, vol. 13, no. 3, pp. 1154–1163, 2016.

[2] J. Teizer, B. Allread, C. Fullerton, and J. Hinze, “Autonomous proactive real-time construction worker safety,” Automation in Construction, vol. 27, pp. 193–202, 2015.

[3] Y. Li, X. Chen, and Z. Wang, “IoT-based wearable safety helmet for hazardous environment monitoring,” IEEE Access, vol. 7, pp. 171600–171610, 2019.

[4] J. Kim and S. Park, “Development of smart helmet system for construction safety,” IEEE Sensors Journal, vol. 20, no. 14, pp. 8001–8009, 2020.

[5] K. Mehata et al., "IoT-Based Safety and Health Monitoring for Construction Workers," Proceedings of the IEEE International Conference on Internet of Things Technology and Communication Engineering, pp. 1-8, 2019.

[6] A.Venkata Subbarao Koduri et al., "A Smart Helmet System for Predictive Maintenance and Safety Monitoring," Proceedings of the IEEE International Conference on Advanced Computing and Communications, pp. 1-8, 2025.

[7] G. Kaur, V. K. Shukla, P. Verma, A. Verma, and C. T. S. Kumar, “Statutory Promise and Classroom Reality: A PRISMA Guided Systematic Review of the Legal Framework for Inclusive Education of Children with Disabilities in India,” International Journal of Special Education, vol. 41, no. 14s, pp. 222–230, 2026.

[8] A. Verma, P. Verma, N. Mansoura, and S. Rajan, “Hybrid deep-learning model for resilient tourism enterprises,” Annals of Tourism Research Empirical Insights, vol. 7, no. 1, p. 100220, 2026.

Biographies



Swati Bhangare Assistant Professor in Electronics & Telecommunication Engineering (E&TC) Department, PVG's College of Engineering, Nashik.



Arati Pingle Assistant Professor in Electronics & Telecommunication Engineering (E&TC) Department. Specialization: VLSI & Embedded Systems, Wireless Communication, IoT.



Sagar Pangavhane Assistant Professor in Electronics & Telecommunication Engineering (E&TC) Department, PVG's College of Engineering, Nashik.



Mrudul Kulkarni Student of Bachelor of Engineering (BE) in Electronics & Telecommunication Engineering (E&TC).



Vivek Jadhav Student of Bachelor of Engineering (BE) in Electronics & Telecommunication Engineering (E&TC).



Nitish Pale Student of Bachelor of Engineering (BE) in Electronics & Telecommunication Engineering (E&TC).



Tanaya Patil Student of Bachelor of Engineering (BE) in Electronics & Telecommunication Engineering (E&TC).

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Driver Drowsiness Detection and Alert System Using Eye Tracking

Shital A. Barahate[1] , Yogesh R. Shinde[2] , SAKSHI V. BHUSARE[3], AMRUTA G. DAHALE[4], VAISHNAVI D. GAIKWAD[5]

[1,2]Asst. Prof. Electronics & Telecommunication Engineering Department, PVG's COE & SSDIOM, Nashik Asst. Prof. Engineering Science Department, PVG's COE & SSDIOM, Nashik [3,4,5]UG Students, Electronics & Telecommunication Engineering, PVG's COE & SSDIOM, Nashik,
[1]shital.barahate24@gmail.com, [2]yogesh.shinde@pvgcoenashik.org,
[3]bhusaresakshi1106@gmail.com,
[4]amrutadahale761@gmail.com, [5]vaishnavig602@gmail.com

Abstract

Driver drowsiness is one of the leading causes of road accidents worldwide, leading to reduced alertness, poor decision-making, and slower reaction time. This paper presents an Eye Tracking-Based Driver Drowsiness Monitoring and Alert System designed to detect early signs of drowsiness using real-time Image Processing Techniques. The proposed system observes blink rate, eye movements, and eye closure duration using an eye-blinking sensor. Key parameters such as Percentage of Eye Closure (PERCLOS) and Eye Aspect Ratio (EAR) are used to determine the fatigue level. When fatigue is detected beyond a predefined threshold, an alert system is activated to alert the driver. The system is cost-effective, non-intrusive, and suitable for real- time implementation in Smart Vehicles as well as Advanced Driver Assistance Systems (ADAS). When a driver is demonstrating signs of fatigue, a buzzer is switched on to alert both the driver and to avoid crashes. Experimental tests show the detection reliability in varied lighting, enabling it to be used with real-world vehicular lighting.

Keywords: Drowsiness Detection, Driver Fatigue, Eye Blinking Sensor, Eye Tracking, EAR, PERCLOS, Road Safety.

1. INTRODUCTION

Drowsy driving is a critical issue in today's transportation system. Continuous focus on the road, long driving hours and lack of sleep, considerably reduce drive alertness. The traditional methods to detect drowsiness are physiological signs including heart rate, EEG monitoring, steering behavior analysis and tracking lane deviation, which are costly and intrusive. However, they have some disadvantages such as high implementation cost; contact dependence with sensors; inability to detect early fatigue [2]. Whereas, vision based systems provide a nonintrusive practical solution. Real time monitoring of driver alertness [1] is an area where eye tracking techniques can be used to analyze. The proposed Eye Tracking Based Driver Drowsiness Detection and Alert System makes use of an eye blinking sensor which captures eye motion [4]. If any of the parameters is measured above the set limits, an alarm mechanism is activated (buzzer, audio alarm) to draw the driver's attention [2].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

It provides real-time monitoring with Low latency, which makes it highly suitable for long-distance transport, commercial vehicles and smart automotive systems. Real time processing is an important component in accident prevention, ensuring timely action. Studies denote that drowsy driving is one of the primary causes of traffic related injuries worldwide [3]. Automotive systems now include Driver Monitoring Systems (DMS) that continuously monitor driver vigilance and make the system safer [5]. Such systems monitor behavior as it happens and warn drivers when certain indicators of driver carelessness are detected. Automotive manufacturers include them in their vehicles to make sure advanced features are used safely [6].

2. RELATED WORK

Table. 1. This work is compared with related studies and previous research by other researchers.

Reference	Eye Tracking	Blink Rate Analysis	PERCLOS	Real Time Alert
Deshmukh P., Kulkarni P. (2022)	Yes	Yes	No	Yes
Asghar R., Hassan A., Junejo N. U. R., Ikram	Yes	Yes	No	Yes
Sun W., Wang Y., Hu B., Wang Q. (2024)	Yes	No	Yes	No
Li, Wan-ning (2024)	Yes	No	No	No
This work	Yes	Yes	Yes	Yes

2.2 Overall Workflow of the Eye Tracking-Based Driver Fatigue Monitoring and Warning System

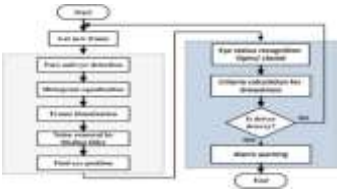


Figure2.2.1 Workflow of the Eye Tracking Based



Figure2.2.2 Importance of DMS

3. METHODOLOGY

4.1 System Architecture

The proposed system consists of the following major modules:
1.Camera Based Eye Tracking Module 2. Infrared Eye Blink Sensor Module 3. Sensor Fusion Engine 4. Artificial Intelligence Based Drowsiness Classification 5. Driver Alert

and Safety Response Module 6. Cloud Based Data. The proposed method enhances conventional approaches by introducing temporal eye movement behavior modeling, which helps detect cognitive drowsiness before complete eye closure occurs.
hybrid driver drowsiness monitoring framework Logging and Predictive Analytics (Advanced Feature).

4.2 Camera Based Eye Tracking Module

A high definition near infrared camera is mounted on the dashboard to continuously capture the driver's facial region. The proposed methodology is implemented differently from the traditional systems which only detect whether or not the eyes are closed:

- Real time iris localization
- Real time embedded implementation
- Improved nighttime and occlusion

The camera stream is detected using a deep learning based facial landmark detection algorithm. The algorithm adapts its significance of each sensor according to the ecological state.
A Driver Drowsiness Index (DDI) is computed by the fusion engine and is based on the following:

- Eye closure duration
- Blink irregularity
- Gaze deviation
- Head stability
- Sensor confidence score
- In this adaptive fusion method, the number of false fusions is reduced. Stabilises detection and enhances it in a positive way. The eye region is dynamically isolated by means of: adaptive region segmentation to reduce computational overhead.
- The system calculates:
- PERCLOS (Percentage of Eye Closure)
- Blink frequency

4.3 Eye Blink Sensor Integration

Another set of eye blink sensors are the infrared type. This is incorporated into smart glasses and mounted Close to driver's eye area. The sensor continually records the interruption of the eyelids Identifies and produces electrical signals. corresponding to blink activity. As compared to earlier research in which the blink sensors are used independently, this work introduces:

- Adaptive blink threshold calibration
- Real-time blink waveform analysis
- False blink elimination using signal filtering
- Sensor confidence scoring

This blink sensor is used as an extra safety measure. Validation layer for the camera system. Even if the camera stops working properly due to The detector can be damaged if it is exposed to a camera dark current, such as low light, or closed face.). continue monitoring fatigue continually. The use of both methods simultaneously makes the Have a system that is much more reliable than systems those which are based on a single source.

4.4 Multi-Modal Sensor Fusion Mechanism

The proposed approach is novel in terms of the key novelty, proposed Hybrid sensor fusion framework. It fuses information from the camera with information from the blink sensor by synchronizing them in time with time stamp based fusion. The drowsiness detection is then refined with a weighted decision making process to increase the accuracy and consistency of the detection. This multi stage classification provides earlier warning capability compared to binary alert systems.

4.5 AI-Based Drowsiness Prediction Model

The proposed research introduces a deep learning model that would predict drowsiness, rather than just detect it when it is really severe. The method is based on a CNN (Convolutional Neural Network) and LSTM (Long Short Term Memory) network to enhance the detection accuracy. The CNN is applied to obtain the important eye related drowsiness, and the LSTM is applied to analyse the pattern changes of drowsiness over time. The system has been developed by integrating the spatial and temporal analysis to better and earlier detect signs of drowsiness.

4.6 Intelligent Alert Generation System

Once the level of drowsiness goes beyond set limits, the system triggers a multi-level alarm:
Level 1: Visual dashboard notification
Level 2: Audio alarm with adaptive intensity
Level 3: Seat vibration alert
Level 4: Advanced Safety Feature: Vehicle speed control recommendation

4. RESULT

The criteria involved to evaluate the performance of the system are accuracy, precision, recall and response time. The experimental comparisons showed that the proposed model accuracy is almost 90-95% in case of Drowsiness detection. System has also been efficient in real time application as it has very less delay. Some loss of performance in less than optimal lighting or when the front of the user's face was obstructed was detected.

5. CONCLUSIONS

This study introduces an efficient and dependable eye-tracking based system for detecting driver drowsiness.By analyzing eye behavior using parameters such as EAR, the system can accurately identify drowsiness in real time.

6. REFERENCES

[1] Deshmukh P., Kulkarni P. (2022), Driver fatigue detection using eye tracking and ma- chine learning, International Journal of Engi- neering Research Technology (IJERT), 11(4), 250–254.

[2] Safarov F., Akhmedov F., Abdusalomov A. B., Nasimov, R. Cho, Y. I. (2023), Real-Time Deep Learning-Based Drowsiness Detection: Leveraging Computer-Vision and Eye-Blink Analyses for Enhanced Road Safety, Sensors, 23(14), 6459.

[3] Asghar R., Hassan A., Junejo N. U. R., Ikram F., Cheema A. M. (2023, June 28), Drowsiness Detection and Alertness Using Eye Motion Monitoring, Sir Syed University Research Journal of Engineering Technology, 13(1), 101–106.

[4] Sun W., Wang Y., Hu B., Wang Q. (2024),Exploration of Eye Fatigue Detection Features and Algorithm Based on Eye-Tracking Signal, Electronics, 13(10), 1798.

[5] P. Verma and R. Verma, "Predictive analytics for sustainable investment portfolios in banking," Multidisciplinary Science Journal, vol. 7, 2025, Art. no. 2025ss0231.

[6] E. P. Pathrose and P. Verma, "The role of higher education in UAE: Narrowing the gap between employers' expectations and students' perceptions," Electrochemical Society Transactions, vol. 107, no. 1, pp. 4213–4231, 2022.

Biographies



Shital A. Barahate received the bachelor's degree and the master's degree in electronics and telecommunication engineering from Savitribai Phule Pune University.



Yogesh R. Shinde received the bachelor's degree and the master's degree in Electrical engineering from Savitribai Phule Pune University.



Sakshi V. Bhusare UG Students, Electronics & Telecommunication Engineering, PVG's COE & SSDIOM, Nashik



Amruta G. Dahale UG Students, Electronics & Telecommunication Engineering, PVG's COE & SSDIOM, Nashik.



Vaishnavi D. Gaikwad UG Students, Electronics & Telecommunication Engineering, PVG's COE & SSDIOM, Nashik.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Automatic Portfolio Website Builder: A Web-Based System for Automated Portfolio Generation and Skilled-Based Job Recommendation

Prof. V. R. Lele[1], Prof. S. K. Thakare[2], Mr. Prathamesh. R. Patil[3], Mr. Tanmay. S. Landge[4], Mr. Yash. P. Bagul[5,] Mr. Yash. V. Pagar[6]

[1,2]Assistant Professor, Department of Information Technology,PVG's COE & SSD IOM, Nashik, India

[3456]Students, Department of Information Technology,PVG's COE & SSD IOM, Nashik, India

Abstract

Modern recruitment systems still face challenges in creating professional portfolios and identifying suitable jobs based on users' skills. This paper proposes an integrated system that automatically generates dynamic portfolio websites and recommends skill-based jobs through a simple web interface. The recommendation module uses TF-IDF and cosine similarity to match skills with relevant job titles. Experimental results demonstrate a response time of approximately 1–2 seconds and effective job recommendations. The proposed system provides a practical and accessible solution for students and job seekers without requiring technical expertise.

Keywords. Online Portfolio, Skill-Based Job Recommendation, Resume Parsing, Natural Language Processing, Web Application

1. INTRODUCTION

The growth of online portfolios has enabled employers to evaluate candidates based on their qualifications, skills, and professional experience. However, creating a professional portfolio remains challenging for users without web development expertise. Existing job recommendation systems also struggle to accurately match users' skills with job requirements and often lack comprehensive candidate profiles. The separation between portfolio generation and job recommendation further limits the effectiveness of digital recruitment platforms. To address these limitations, this study proposes an integrated online system that automatically generates professional portfolios and recommends suitable jobs based on users' skills. The system provides a user-friendly interface and applies techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) and cosine similarity to match skills with relevant job opportunities, supporting efficient career development and employment guidance.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

Research on portfolio generation and job recommendation has evolved from basic resume builders toward automated and convenient solutions. Early studies explored resume generation and web-based profiles [4]. Template-based portfolio generation was proposed to simplify website creation for non-expert users [2], [5], but these approaches primarily focused on content organization rather than intelligent recommendations. Later studies introduced dynamic template generation to improve flexibility, information presentation, and user experience [3]. Job recommendation systems have also employed skill-matching using similarity-based methods [6] and AI-based approaches [7]. Some studies integrated profile building with recommendation functionality in a single system [1]. However, existing solutions generally treat portfolio generation and job recommendation as separate or loosely connected tasks, limiting the use of structured portfolio information for accurate skill-based recommendations. The proposed system integrates portfolio creation, hosting, management, and job recommendation using TF-IDF and cosine similarity, improving usability, accessibility, recommendation quality, and career opportunities.

3. PROPOSED SYSTEM

The proposed web-based system integrates frontend, backend, database, portfolio generation, recommendation, and deployment modules to support portfolio management and skill-based job recommendations. The backend uses Node.js and Express.js for authentication, validation, and application services, while MySQL stores user profiles, skills, and job information. The portfolio module converts user inputs into a predefined structure and generates customizable portfolio websites that users can preview and edit.

The job recommendation module evaluates user skills and compares them with the skills required for different jobs available in the database. The importance of skills is calculated by using the TF-IDF model:

$$TF - IDF(t,d) = TF(t,d) \times log(N / DF(t)) \quad (1)$$

The similarity between user skills and job requirements is calculated using cosine similarity:

$$Cos(\theta) = (A \cdot B) \times (||A|| \times ||B||) \quad (2)$$

Here, A = User skill vector B = Job skill vector, respectively. Job posting is ranked based on the scores and top results are given to the user.

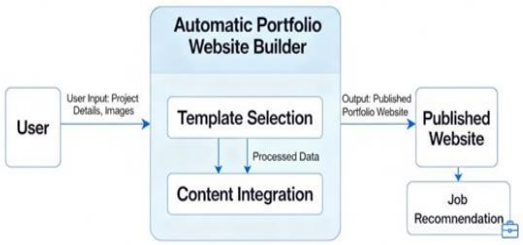


Figure 1. Proposed System Architecture

After portfolio generation, users can choose to publish their portfolio as a live website. The system hosts the portfolio and generates a public link, allowing users to control when it becomes publicly accessible. Figure 1 illustrates the proposed architecture, comprising frontend, backend, database, portfolio generation, and recommendation modules. Users provide profile and project information, after which the system selects a suitable template and generates the portfolio. The recommendation module then suggests relevant job positions based on the user's skills and portfolio information.

4. RESULTS AND DISCUSSION

The system was evaluated using response time, recommendation accuracy, and usability. Response times ranged from 1–2 seconds, indicating efficient operation. TF-IDF with cosine similarity outperformed rule-based and frequency-based methods in Precision, Recall, and F1-score. Recommendation relevance was measured using the proportion of required job skills matched by the user profile. For example, matching three of four required skills produces a 75% matching score, indicating strong job relevance. Usability was assessed through portfolio completion, representing the percentage of completed sections. Completing three of four sections results in a 75% completion rate. Testing showed that the system handled multiple simultaneous requests, updated recommendations immediately after skill changes, and provided a user-friendly interface suitable for both technical and non-technical users.

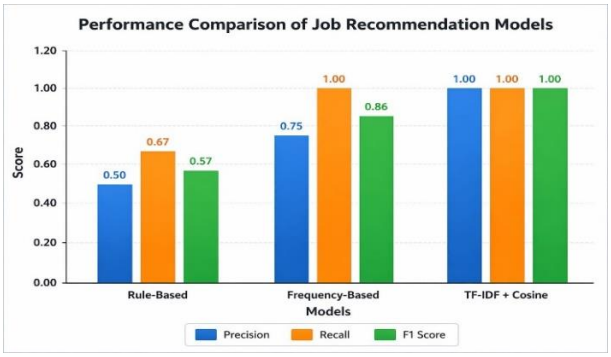


Figure 5.1. Grouped Bar Chart showing performance comparison of recommendation models

5. CONCLUSION

The proposed system enables efficient portfolio creation through a responsive, user-friendly dashboard without requiring technical knowledge. Its job-matching module provides skill-based career recommendations, reducing the effort required to identify suitable opportunities. However, recommendation accuracy depends on user-provided skill data and may be affected by semantic differences. Future work will incorporate advanced machine learning and NLP techniques, along with real-time job data, to improve recommendation accuracy and overall performance. Overall, the system integrates portfolio creation, management, and job recommendations into a centralized platform, saving time and improving accessibility for less technically skilled users.

6. REFERENCES

[1] S. Jadhav, S. Kawade, and S. Chikhale, “ICT: Resume Builder System Using Full Stack Web Development,” *IJRASET*, vol. 11, no. 5, pp. 6912–6920, May 2023, Doi: 10.22214/ijraset.2023.53264.

[2] A. Sharma and R. Patel, “Automated Personal Portfolio Generation Using Web Templates,” International Journal of Research Publication and Reviews, vol. 5, no. 11, pp. 1580–1583, Nov. 2024.

[3] Ashish Pundhir, Himanshu, Kushal Sharma, Gaurav Singh Sirari,” Online Portfolio maker by using a single website”, JETIRJ, vol. 11, May-2024.

[4] N. Pramanik, "Job Recommendation System Based on Candidate Skills," in IEEE, Villupuram, India, 2024. Doi: 10.1109/ICSTSN61422.2024.10671166.

[5] S. Kumar, N. Rajput, V. Mishra, and G. Aggarwal, “Intelligent Hiring Software – Job Recommendation System and Profile Builder,” in Proc. IEEE Int. Conf. Advances in Computer Science, Electrical, Electronics and Communication Technologies (CE2CT), 2025, pp. 1–6, April 2025, doi: 10.1109/CE2CT64011.2025.10941622.

[6] Sudeep MA, Mahalakshmi M, “Dynamic Template Designing Web Application “, IJPREMS, Vol. 05, October 2025.

[7] G. Yogeswara Rao et al., “AI-Based Job Recommendation System Using Skill Extract & Matching,” IJRASET, vol. 14, no. 3, Mar. 2026 Doi:10.22214/ijraset.2026.77976.

Biographies



Prof. Vaishali R. Lele received the bachelor's degree in Electronics from Shivaji University, Kolhapur, Maharashtra in 1994, the Master’s degree in VLSI & Embedded Systems from SPPU, Pune in 2013. She is currently working as an Assistant Professor at the Department of Information Technology, PVG’s College of Engineering & S. S. Dhamankar Institute of Management, Nashik. She has been served as a resource person for faculty development programmes under SPPU, reviewer for many highly-respected journals, etc.



Prof. Smita K. Thakare received the bachelor's degree in Computer Engineering from Savitribai Phule Pune University, the Master’s degree in Computer Engineering from SPPU, Pune. She is currently pursuing Ph.D. in computer engineering and science from Pimpari

Chinchwad University, Pune. She is currently working as an Assistant Professor at the Department of Information Technology, PVG's College of Engineering & S. S. Dhamankar Institute of Management, Nashik. She has total 12 years of experience.



Prathamesh Rajendra Patil received his Bachelor of Engineering degree in Information Technology from Pune Vidyarthi Griha's College of Engineering, Nashik, in 2026. His areas of interest include web development and software development.



Tanmay Landge received his degree in Information Technology from Pune Vidyarthi Griha's College of Engineering, Nashik, Savitribai Phule Pune University in 2026 and specialize in web development. Alongside his technical work, he founded Blink Link, a venture focused on helping businesses strengthen their digital presence through technology-driven solutions. With a strong interest in technology and entrepreneurship, his goal is to create practical solutions that help businesses grow.



Yash Prakash Bagul received his degree in Information Technology from Pune Vidyarthi Griha's College of Engineering, Nashik, Savitribai Phule Pune University in 2026. His area of interests includes programming and development in python, AI, data science, Machine learning.



Yash Pagar received his Bachelor of Engineering degree in Information Technology from Pune Vidyarthi Griha's College of Engineering, Nashik, in 2026. His areas of interest include software development, data science and machine learning.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Design and Implementation of a Decentralized Uptime Monitoring System with Byzantine Fault Tolerance on Solana

Satish Bhadane[1], Dipali Shinde[2], Dhananjay Ahire[3], Ritesh Chaudhari[4], Sujit Gaware[5], Chetan Chavan[6]

[1]*Asst. Prof. Information Technology, PVGs COE, Nashik, satish.bhadane@gmail.com*
[2]*Asso. Prof., Information Technology, PVGs COE, Nashik, da.shinde2709 @gmail.com*
[3] *Information Technology, PVGCOE-SSDIOM, Nashik, dhananjayahire15@gmail.com*
[4] *Information Technology, PVGCOE-SSDIOM, Nashik, riteshchaudhari7100@gmail.com*
[5] *Information Technology, PVGCOE-SSDIOM, Nashik, sujitgaware55@gmail.com*
[6] *Information Technology, PVGCOE-SSDIOM, Nashik, chetanschavan3000@gmail.com*

Abstract.

Uptime monitoring platforms are like watchdogs for any websites or services. that make sure their online presence is there for their users. A minute of downtime can cost companies a large bite of their revenue. Uptime monitoring platforms make that any service it is monitoring is online, and the second it stops responding, it sends Notification to the service owner—a great way to monitor your online presence, isn't it? Yes Of course. But the issue many uptime monitoring platforms face today is their A centralized architecture, which is vulnerable to attacks, depends heavily on cloud infra. and does not really cover every corner of the world. Typically a centralized uptime monitoring platform will have a cloud server hosted in another part of the world from which it monitors any service. A server in a big city of a country doesn't represent the whole country's status. The solution to this problem is decentralization. Building an uptime monitoring platform on top of Depin (Decentralized Physical Infrastructure Network). DePIN is technology. This allows us to decentralize the physical and infrastructure layer of any system. In This way, the centralized infrastructure model of traditional monitoring systems can be distributed to many worker nodes who are just individuals with compute power. Individually acting as a server that monitors a service. Decentralizing this system It comes with some challenges, like how to trust our worker nodes? How to calculate the result? How to incentivize them?. To tackle this problem, the Solana block chain comes into the picture. Allowing us to implement a consensus algorithm for deriving the final result, smart contracts to disincentivize worker nodes and Solana as a crypto to reward the worker nodes. This decentralized uptime monitoring system can allow services to monitor their website. From every corner of the world, and trust the results, as they are not produced by a single entity, which a centralized version lacks.

Keywords. Decentralized monitoring; Consensus algorithm; Solana block chain; Distributed validators; Uptime verification.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

A website might go down form time to time its API endpoints might go down from time to time. This can cost the website owners big chunk of their revenue and lower customer satisfaction. To avoid this many websites use Uptime monitoring platforms to monitor the online presence of their website. But this cen-tralized uptime monitoring systems can't cover every corner of world as they are hosted on cloud servers in a city. The results of this platform are also not tam-per proof which leads in high rates in false results. Traditional monitoring platforms like Pingdom, UP time robot, better uptime and many more faces these issues. The solution to tackle this problem is to use Depin on top of Uptime Monitoring platform. De-pin is Decentralized Physical Infrastructure network which uses the blockchain technology to distribute the creation of real world infrastructure [7]. By using De-pin we can decentralize the cloud server of centralized monitoring platforms into various small servers. This small servers or working nodes are just people around the globe with compute power monitoring a service. These worker nodes are called validators. Multiple validators from a region will monitor a website and will come to conclusion about a website's status [2]. A region can have multiple validators and there may be conflict between validators in agreeing upon on same result. To solve this we can introduce a consensus algorithm [5] which handle all validators result and produce final result. Honesty of validator is also a big concern in this model. Smart contracts can be used to implement reward and punish mechanisms for validators [1]. As the results are not produced by single entity but by many validators.

2. BACKGROUND AND RELATED WORK

2.1 Traditional Centralized Monitoring Systems

Traditional uptime monitoring systems mainly depend on centralized cloud servers placed in a few major locations around the world. These systems monitor website availability, response time, and server status continuously.

2.2 Blockchain-Based Decentralized Monitoring

The idea of decentralized monitoring gained attention after the introduction of blockchain-based applications by V. Buterin [1] through Ethereum smart contracts. This approach showed that monitoring results could be stored and verified in a transparent and tamper-resistant manner without relying on a single authority. Later, Zhang and Patel [3] proposed a distributed monitoring framework where multiple monitoring probes collected uptime data and validated results using blockchain consensus. Sharma et al. [4] further discussed decentralized uptime verification using multi-source monitoring and trust-based validation mechanisms to reduce false alarms and improve reliability.

2.3 Research Gap

Ubiwatch unifies distributed validators, WebSocket communication, blockchain verification, decentralized payments, and global monitoring for scalable, reliable, real-time website uptime monitoring.

3. METHODOLOGY

3.1 Overview of the Proposed Framework

Decentralized uptime monitoring uses Solana incentives, Byzantine fault tolerance, distributed validators, database integration, centralized hub coordination, and real-time website status verification.

3.2 Implementation flow

This procedure comprises user setup, validator start-up, periodic checking, signing proofs, consensus validation, Solana logging, reward calculation, and WebSocket alerts through dashboard updates.

3.2 Architecture Explanation

Figure 1 shows an architecture of a decentralized monitoring mechanism using blockchain technology that includes wallet authentication, SOL payment, validators, real-time WebSockets communication, results aggregation, data storage, and uptime visualization.



Figure 1. Glacier Segmentation Architecture

4. EXPERIMENT RESULTS

4.1 Experimental Setup

A controlled web service was deployed locally and exposed using secure tunneling to simulate real-world conditions.

4.2 Byzantine Fault Tolerance Results

Table 1 BFT Testing Results

Byzantine Nodes	Consensus	Correctness	Theory
0 (0%)	Yes	100%	Pass
1 (14.3%)	Yes	100%	Pass
2 (28.6%)	Yes	100%	Pass
3 (42.9%)	No	N/A	Fail

4.3 Results and Analysis

Metric	DecentralWatch	Pingdom	Datadog	Uptime Kuma	UptimeRobot
Accuracy	**99.92%**	99.80%	99.70%	98.50%	99.30%
F1-Score	**0.9921**	0.9854	0.9850	0.9648	0.9831
False Positive Rate	**0.81%**	2.10%	1.80%	3.20%	2.80%
Consensus Latency	2.3s	50ms	80ms	N/A	60ms
Alert Latency	1.2s	300ms	400ms	2.1s	400ms
Cost/Month	**Free**	$15+	$15+	Free	$9.99
Byzantine Tolerance	**Yes** ($f = 2$)	No	No	No	No
Decentralization	**Full**	None	None	Partial	None

Figure 2: Ubi-Watch vs Centralized Systems (30-day evaluation)

Figure 2 shows Decentral-Watch achieves higher accuracy, fewer false positives, Byzantine tolerance, decentralization, transparency, reliability, and cost-effective real-time monitoring.



Figure 3: Validator Dashboard

Figure 3 shows the validator dashboard. It shows status (Active/Inactive), pending payouts in Solana, and how many website validators have monitored it and their average uptime.

5. CONCLUSION

The proposed system uses a decentralized network where multiple validators verify uptime and reach agreement through blockchain-based consensus. This approach improves transparency, reliability, fault tolerance, and trust by distributing control. Although consensus may introduce slight delays, the system reduces errors and strengthens availability. Future improvements can focus on faster consensus, scalability, and efficient validator coordination.

6. REFERENCES

[1] V. Buterin, “A Next-Generation Smart Contract and Decentral-ized Application Platform,” Ethereum Whitepaper, 2015.

[2] M. K. Jayanthi Kannan, A. Yadav, V. Kumar, N. H. Pramod, A. Tomer, and F. A. Khan, “SiteWatch: Uptime Tracking and Management System for Real-time Website Health,” Journal of Emerging Technologies and Innovative Research (JETIR), vol. 11, no. 12, 2024.

[3] M. Zhang and D. Patel, “Blockchain-Based Distributed Monitor-ing Framework for Cloud Services,” International Journal of Computer Applications, vol. 179, no. 32, pp. 22–29, 2022.

[4] P. Sharma, L. Gupta, and T. Singh, “Improving Web Reliability through Decentralized Uptime Verification,” ACM Comput-ing Surveys, vol. 55, no. 7, pp. 1–20, 2023.

[5] P. Verma, R. Jain, and A. Bansal, “Consensus Mechanisms for Decentralized Reliability Monitoring,” Journal of Network Systems and Management, vol. 31, no. 1, 2024.

[6] J. Li and H. Chen, “Design of Decentralized Monitoring Archi-tecture Using Dockerized Validator Nodes,” IEEE Transac-tions on Cloud Computing, vol. 12, no. 3, pp. 514–523, 2024.

[7] R. B. Dos Passos, K. J. Matteussi, J. C. S. Dos Anjos, and C. F.

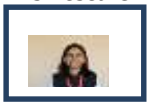
[8] R. Geyer, “Towards a Decentralized Blockchain-Based Re-source Monitoring Solution for Distributed Environments,” Journal of Internet Services and Applications, vol. 15, no. 1, 2024, doi: 10.5753/jisa.2024.3813.

[9] M. Oliveira and P. Costa, “Integrating Blockchain and IPFS for Transparent Service Monitoring,” IEEE Transactions on Emerging Topics in Computing, vol. 13, no. 1, pp. 295–308, 2025.

Biographies



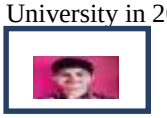
Satish Bhadane is an academician and researcher in the field of Computer Engineering**,** His research work has received 312 citations, an h-index of 9, and an i10-index of 5**.**



Dipali Shinde is an academician and researcher in the field of Computer Engineering**;** she has published several research papers in international journals and conferences.



Dhanjay Ahire received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.



Ritesh Chaudhari received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.



Sujit Gaware received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.



Chetan Chavan received the bachelor's degree in computer engineering from SPPU Pune, University in 2026.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An AI-Powered Predictive Cloud Resource Allocator Using Prometheus-Based Workload Telemetry

Dr. Chetan H. Patil[1], Prof. Sanket G. Chordiya[2], Prof. Hansa S. Borse[3], Atharva V. Aher[4], Akshada B. Dinde[5], Amit S. Chaudhari[6], Vaishnavi V. Darekar[7]

Assistant Professor,chetanhpatil@gmail.com,sanket.chordiya@gmail.com, hansashimpi@gmail.com,Research Scholar, atharva0276@gmail.com, dindeakshu1210@gmail.com, amitchaudhary6399210@gmail.com, vaishnavi.v.darekar@gmail.com,Department of Artificial Intelligence and Data Science, PVG's COE & SSDIOM, Nashik, India

Abstract.

With the introduction of cloud computing, there has been an intense revolution in the digital infrastructure space by offering on-demand and scalable computing resources. Conventional fixed or deterministic resource allocation policies, on the other hand, often lead to either over-allocation or under-utilization and thus trigger a decrease in performance or unnecessary expenditures. This paper presents a framework based on AI-powered Cloud Resource Allocator and Manager, also known as the proposed framework, that uses the models of Machine Learning (ML) and Artificial Intelligence (AI) to predict workload fluctuations, identify anomalies, and automatically coordinate the decisions on cloud scaling. The framework uses time-series forecasting models, including Long Short-Term Memory (LSTM) [1] and eXtreme Gradient Boosting (XGBoost) [2] models with the support of anomaly refinement methods to achieve an effective balance between costs and performance [3]. The proposed model can be used to achieve effective adaptive scaling, better anomaly refinement and an improved cost utilization, according to the tests based on simulated AWS EC2 instances and the monitoring dataset [4]. The obtained results show that pro- active cloud resource allocation can effectively increase the reliability and cost-effectiveness compared to the traditional methods. [5], [6].

Keywords. Artificial Intelligence; Cloud Computing; Machine Learning; Predictive Scaling; Resource Optimization; Anomaly Refinement.

1. INTRODUCTION

Cloud environments present a major hurdle as they require resources to be used efficiently on the one hand and performance of the cloud environment to not degrade on the other; and minimize cloud operational cost. The traditional methods of allocating resources are based on manual configuration or statically configured resource scaling (artificially configured rules) mechanisms that provision resources only when pre-defined thresholds are surpassed. These reactive solutions may result in postponing responses, impairment of performance, and needless waste of resources. The limitations are overcome by the advancements of Artificial Intelligence (AI) and Machine Learning (ML) with their "proactive" (data driven)

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

and “self-adaptive” resource allocation methods. Using past usage trends and real-time performance metrics, ML models can forecast workload patterns and make automatic scaling decisions before hitting bottlenecks. This predictive resource management results in an optimal cost/performance balance. This paper proposes an AI-Powered Cloud Resource Allocator and Manager that integrates predictive modeling, anomaly refinement, and optimization-based decision-making.

2. SYSTEM ANALYSIS

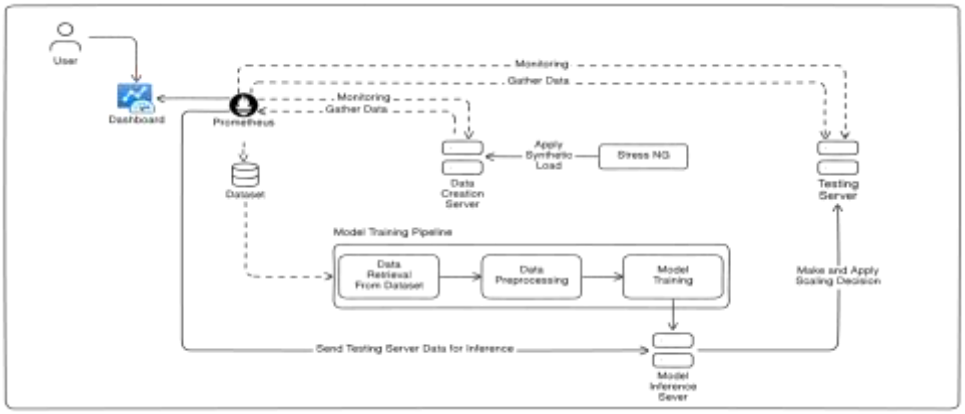


Figure 2.1. System Architecture and Workflow

The suggested AI-Powered Cloud Resource Allocator aims to maximize cloud infrastructure utilization by forecasting future resource demands and automatically adapting system configurations. This section presents both the implemented phase and the projected behavior of the complete system to analyze its feasibility and performance.

2.1. Telemetry Collection Module

This module is developed in order to collect real-time metrics from the server being monitored. Prometheus is used to monitor, while node exporter exposes the low-level system statistics. Prometheus scrapes the metrics at a short time interval, generating time-series data. This data acts as the primary input to the predictive system[7],[8].

2.2. Predictive Machine Learning Module

The Predictive Machine Learning Module is the intelligent component of the system, responsible for predicting the CPU Utilization at the next 60 seconds mark (T + 60) using the currently gathered telemetry data. The XGBoost predictive model is the primary deployed model, having the capabilities of capturing non-linear patterns in the workload. The predicted CPU metric is then forwarded to the decision engine to perform the scaling actions [2],[7].

Table 2.1. Comparative Analysis

Method	Scaling Type	Test RMSE	Remarks
Osypanska and Nawrocki (2023)	ML-Based	41.18	Low Precision

Proposed System	Proactive	22.26	Fully Automated Closed Loop

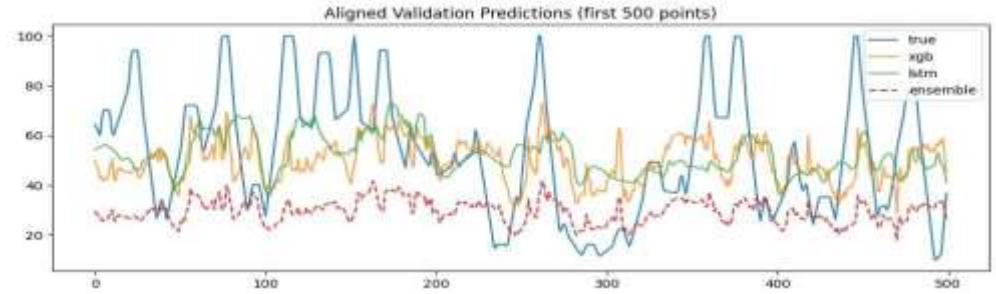


Figure 2.2. Implementation Result

2.3. Scaling Decision Engine

The Decision Engine processes the CPU usage predicted by the Machine Learning (XGBoost) model. It then determined whether the server should be scaled or not. It operates based on the predefined parameters such as thresholds for CPU utilization, model confidence and cool down period to prevent instability and frequent scaling. The module acts as the core, converting the insights into instructions [9],[10].

2.4. Cloud Infrastructure Control Module

Once the scaling is required the Cloud Infrastructure Control Module communicates with the cloud provider's APIs to adjust the resources that are used by the running instance.This process allows vertical scaling where you can scale up and scale down the computational capacity of the instance as needed, based on the predicted workload demands. Due to automation, there is no need for manual operations, and the infrastructure can be reconfigured dynamically and responsively depending on the expected load on the system[11],[12].

2.5. Fully Automatic Loop (Autoscaler Daemon)

The entire system operates through a continuously running autoscaler daemon that coordinates the interactions between all system modules. The daemon periodically performs the following sequence of operations: 1. Query telemetry data from Prometheus. 2. Generate predictive features from the collected data. 3. Perform machine learning inference to forecast CPU utilization. 4. Evaluate scaling policies based on predicted workload conditions. 5. Scale the infrastructure when necessary. 6. Reinitialize monitoring on the server after successful scaling. This allows a closed-loop system to manage a cloud server autonomously, without needing human intervention [5],[13],[12].

3. CONCLUSION

The proposed system shows that Machine Learning can be effectively used to optimize cloud resource utilization. It ensures efficient allocation and dynamic adjustment of computing resources while maintaining performance and minimizing downtime. Using tools such as Prometheus, Node Exporter and Stress-NG can help to gather a representative, complete and realistic set of data. Using the models LSTM and XGBoost, patterns can be detected in the data of the servers for accurate predictions. It is found that the system can achieve high

accuracy and adaptability in various scenarios through experimental results. Integration with Cloud APIs allows direct implementation of model predictions for real-time optimization.

4. REFERENCES

[1] S. Hochreiter and J.Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[2] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785– 794.

[3] Y. Zhang, X. Chen, and L. Wang, "A Hybrid Anomaly Detection Framework for Cost–Performance Optimization in Time-Series Forecasting," IEEE Access, vol. 9, pp. 145321–145333, 2021.

[4] A. Gupta, P. Verma, and S. Nair, "Machine Learning-Based Cloud Resource Optimization Using Predictive Scaling," IEEE Transactions on Cloud Computing, 2023.

[5] R. Buyya, M. Yousif, and S. Pandey, "Intelligent Cloud Resource Manage ment: Foundations and Future Directions," Journal of Cloud Computing, 2022

[6] L. Xu, J. Rao, and X. Bu, "A Reinforcement Learning Approach to Cloud Resource Management for Elasticity and Cost Efficiency," IEEE Transactions on Network and Service Management, vol. 19, no. 1, pp. 11–24, 2022.

[7] S. Kumar, A. Singh, and R.Buyya, "Workload Forecasting in Cloud Data Centers Using Machine Learning: A Comparative Study," IEEE Transactions on Cloud Computing, vol. 10, no. 5, pp. 3201–3215, 2022

[8] J. Wu, Q. Deng, and H. Chen, "Workload Prediction in Cloud Computing Using Hybrid Deep Learning Model," IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 4, pp. 957–970, 2020.

[9] S. Wang, R.Buyya, and J. Chen, "Energy-Efficient Load Balancing and Auto-Scaling in Cloud Data Centers Using Reinforcement Learning," IEEE Transactions on Cloud Computing, vol. 9, no. 3, pp. 1080–1092, 2021.

[10] A. Beloglazov and R Buyya, "Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers," Concurrency and Computation: Practice and Experience, vol. 24, no. 13, pp. 1397–1420, 2012.

[11] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," NIST Special Publication 800-145, National Institute of Standards and Technology, Gaithersburg, MD, USA, 2011.

[12] M. A. Khan, S. U. Rehman, and M. U. Ashraf, "Energy and Cost-efficient Resource Management in Cloud Data Centers Using Predictive Analytics," Future Generation Computer Systems, vol. 108, pp. 803–811, 2020.

[13] J. Gao, H. Shen, and C. Li, "AI-Driven Cloud Resource Management: A Survey of Techniques and Trends," IEEE Transactions on Cloud Computing, vol. 11, no. 3, pp. 1458–1475, 2023.

Biographies



Dr. Chetan H. Patil has a PhD in Computer Science and Engineering and is working as an Assistant Professor in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik. His academic and professional interests include Data Science, Machine Learning & Cyber Security.



Prof. Sanket G. Chordiya is Pursuing a PhD in Computer Science and Engineering and is working as a Head of Department in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik. His academic and professional interests include Database Management Systems, Artificial Intelligence, Data Science, Machine Learning & Cyber Security.



Prof. Hansa S. Borse has completed M.E in Computer Engineering and is working as an Assistant Professor in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik. His academic and professional interests include Data Structures, Artificial Intelligence, Data Science, Machine Learning.



Atharva Aher is a Student in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik



Akshada Dinde is a Student in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik



Amit Chaudhari is a Student in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik



Vaishnavi Darekar is a Student in Department of Artificial Intelligence & Data Science at PVGCOE & SSDIOM, Nashik

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

LearnCraft VR: A Gamified Virtual Reality and AI-Driven Approach for Inclusive Education

Dr. S. P. Khedkar[1], Abhishek Borate[2], Anuj Jadhav[3], Priyanshu Shende[4], Rahul Patil[5]

[1,2,3,4,5]*Department of Computer Engineering, Modern Education Society's Wadia College of Engineering, Pune, India.*

[1]*shilpa.khedkar@mescoepune.org,* [2]*abhishekborate56@gmail.com,*
[3]*anujjadhav170@gmail.com,* [4]*priyanshushende2004@gmail.com,*
[5]*rmpatiljal2004@gmail.com*

Abstract.

Children with learning disabilities, such as dyslexia, dysgraphia, and dyscalculia, often find it hard to cope in traditional learning environments with no adaptive feedback. However, with new technological advancements in the form of VR and AI, there are now unique possibilities to build immersive learning systems for people with learning disabilities. The purpose of this study is to conduct a review of existing studies that investigated the effectiveness of different VR applications for learning, diagnosis, personalized feedback, and gamification. In addition to reviewing different applications of VR technologies for improving phonemic awareness, learning hand writing, developing numerical logic, and increasing students' engagement in class, this research will address the effectiveness of speech recognition software, handwriting assessment programs, and adaptive algorithms as tools for automatically giving learners feedback and measuring performance metrics. Based on existing weaknesses of current VR and AI systems, a brief description of the framework, called LearnCraft VR, designed for creating a unified system for building an immersive environment using VR and AI technologies, will be given.

Keywords. Virtual Reality; Learning Disabilities; Dyslexia; Dysgraphia; Dyscalculia; Gamified Learning.

1. INTRODUCTION

Learning disabilities such as dyslexia, dysgraphia, and dyscalculia can be rather common in adolescents and can negatively affect their learning process if not diagnosed in time. Dyslexia affects language acquisition and phonics, dysgraphia affects writing skills, and dyscalculia affects arithmetical skills. These disorders often go hand in hand and can become the obstacle to proper literacy skills development. Traditional class learning is based on the development of reading and writing skills which require visual perception and are not suitable for every individual. Multisensory teaching and continuous help can be very beneficial for students with SLDs [1, 2].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

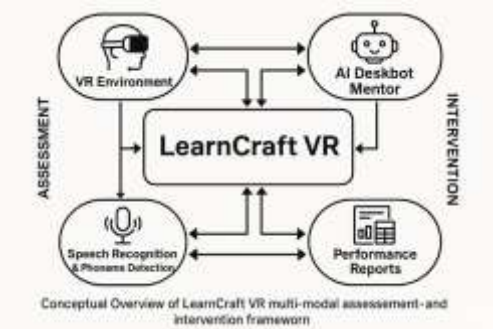
Computing technologies have allowed the creation of the software for dyslexic readers [3, 4] and the computer systems that diagnose dysgraphia [5, 6]. Although highly flexible and efficient, all these applications are made specifically for a particular disorder and the development of a particular skill only. Computerized learning systems with exercises for dyslexia and dyscalculia have been developed [7, 8], but there are no platforms that work in several dimensions and can evaluate the learner's performance.Not much work has been done in the direction of developing a platform that combines all three elements - gamification, virtual reality, and artificial intelligence.

2. RELATED WORK

Research into SLDs via technology has progressed from standard clinical assessments to the application of advanced technological techniques using artificial intelligence, video games, and immersive media.Research into dyslexia showed how it occurred due to issues in phonological abilities and brain functioning [10, 11]. Rello et al. suggested guidelines for website structure design to reduce the cognitive load on people with dyslexia [3, 4]. Even though these developments contributed greatly, they were limited to two-dimensional textual interfaces without live adaptation and feedback.Research into dysgraphia was done in connection to kinematics and writer's movements. Rosenblum et al. noticed the difference between process and product in developmental dysgraphia [12]. Next steps involved classification of handwriting based on the analysis of stroke segments to identify abnormalities [5, 6]. In a similar manner, the research of Geary [1] and Butterworth [13] paved the path for computerized screenings and remediation software [14, 15] in dyscalculia.Gamification has also been used to make practice interesting. Yalu was designed by Kularathna et al. for dyslexia screening [7], and it has been found that motivation-based user interfaces encourage the reading time [8, 16]. But in these approaches, there is no assessment of efficiency in terms of error metrics. In the field of immersive media, VR-assisted cognitive training shows higher attention and motivation in case of problems in reading [9]. But in most cases, the research is limited to prototype development in the individual areas only.

3. PROPOSED FRAMEWORK

The LearnCraft VR system is designed as an integrated solution that incorporates immersive learning environments, game-based learning activities, and AI-based learning analytics. The architecture includes three major components: immersive learning modules, AI analysis modules, and performance monitoring modules.



Conceptual Overview of LearnCraft VR multi-modal assessment-and intervention frameworn

Figure 3.1. Conceptual overview of the LearnCraft VR multi-modal assessment and intervention framework.

4. COMPARATIVE ANALYSIS

Table 1 Comparison of existing approaches with the proposed LearnCraft VR framework

Approach	**AI-based Analysis**	**Gamification**	**Immersive VR**	**Multi-skill Learning**
Dyslexia web tools [3, 4]	No	No	No	Reading only
Handwriting analysis [5, 6]	Yes	No	No	Writing only
Dyscalculia tools [14, 15]	Limited	No	No	Mathematics only
Gamified dyslexia systems [7, 8]	Limited	Yes	No	Reading skills
VR learning systems [9]	Limited	Yes	Yes	Limited scope
LearnCraft VR (proposed)	Yes	Yes	Yes	Reading, Writing, Mathematics

5. DISCUSSIONS

According to the existing literature, while the use of technology for learning has many benefits, the efficiency of the system is restricted by its fragmentary nature. Each tool is designed specifically for one aspect of learning, forcing teachers to utilize different tools. The gamification process increases the level of student engagement [7, 8]; however, the platform is set for working within rigid task constraints. Technologies of automation such as handwriting and speech recognition decrease the necessity of manual evaluation [5, 6], but there are limitations in their application.Immersive VR environments offer structured learning with multi-sensory experience in terms of visual, auditory, and motion, which helps maintain student attention [9, 18]; nonetheless, many VR applications work independently without data analytics and adaptive learning systems. LearnCraft VR provides a solution by combining the elements of VR, machine learning, and gamified learning into one adaptive environment.

6. CONCLUSIONS

Innovations such as digital learning systems, AI-enabled diagnostics, gamification of learning systems, and virtual reality environments can be used effectively to enhance the learning process for dyslexic, dysgraphic, and dyscalculic children. Although many applications are available in an individual form, they tend to act separately and not as an integrated learning system.LearnCraft VR system fills this gap by integrating VR technologies, gaming techniques, and learning analytics via AI within one learning environment. The inclusion of speech recognition and handwriting analysis features, along with automatic assessment of results, makes it useful for enhancing learning processes as well as identifying issues at an early stage. Further research will aim at conducting empirical studies in classroom settings.

7. REFERENCES

[1] D. C. Geary, "Dyscalculia at an early age," Encyclopedia on Early Childhood Development,pp. 1–4, 2006.

[2] N. E. Syah et al., "Computer play pedagogy intervention for children with low conceptual understanding in mathematics," Interactive Learning Environments, vol. 24, no. 7, pp. 1477–1496, 2016.

[3] L. Rello, G. Kanvinde, and R. Baeza-Yates, "Layout guidelines for web text and a web service to improve accessibility for dyslexics," in Proc. Int. Cross-Disciplinary Conf. Web Accessibility, 2012, p. 36.

[4] L. Rello, H. Saggion, R. Baeza-Yates, and E. Graells, "Graphical schemes may improve readability but not understandability for people with dyslexia," in Proc. First Workshop Predicting and Improving Text Readability, 2012, pp. 25–32.

[5] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia using handwriting classifi-cation," IEEE Trans. Human-Machine Systems, vol. 47, no. 2, pp. 293–298, 2016.

[6] J. Mekyska et al., "Identification and rating of developmental dysgraphia by handwriting analysis," IEEE Trans. Human-Machine Systems, vol. 47, no. 2, pp. 235–248, 2016.

[7] T. C. P. Kularathna et al., "Yalu: Computer game based solution to screen learning disabilities in kids," in Proc. IEEE Int. Conf., 2014.

[8] D. Gooch, A. Vasalou, and L. Benton, "Exploring the use of a gamification platform to support students with dyslexia," in Proc. IEEE Int. Conf. Intelligent Systems and Applications (IISA), 2015, pp. 1–6.

[9] J. Lukes, "Dyslexia friendly reader: Prototype, designs, and exploratory study," in Proc. IEEE Int. Conf. Intelligent Systems and Applications (IISA), 2015, pp. 1–6.

[10] R. J. Schain, Neurology of Childhood Learning Disorders. Baltimore, MD, USA: Williams & Wilkins, 1977.

[11] J. G. Guardiola, "The evolution of research on dyslexia," Anuario de Psicolog´ıa, vol. 32, no. 1, pp. 3–30, 2001.

[12] S. Rosenblum, P. L. Weiss, and S. Parush, "Handwriting evaluation for developmental dysgraphia: Process versus product," Reading and Writing, vol. 17, no. 5, pp. 433–458, 2004.

[13] B. Butterworth, "Developmental dyscalculia," in Neuropsychology: Concepts, Theory, and Practice. London, U.K.: Psychology Press, 2008, pp. 357–374.

[14] N. C. Zygouris et al., “Screening for disorders of mathematics via a web application,” in Proc. IEEE EDUCON, 2017, pp. 502–507.
[15] S. Purohit and M. S. M., “Significance of data mining techniques in classifying dyscalculia,” Int. J. Eng. Sci. Comput., vol. 5, no. 8, 2018.
[16] N. Nosworthy, S. Zheng, and D. Ansari, “Kindergarten children’s number comparison skills predict later math scores,” Developmental Science, 2014.

Biographies



Dr. S. P. Khedkar received her Ph.D. in Computer Engineering and is cur-rently working as a Professor in the Department of Computer Engineering at Modern Education Society’s Wadia College of Engineering, Pune, India.



Abhishek Borate completed his Bachelor of Engineering degree in Computer Engineering from Modern Education Society’s Wadia College of Engineering, Pune.



Anuj Jadhav completed his Bachelor of Engineering degree in Computer Engineering from Modern Education Society’s Wadia College of Engineering, Pune.



Priyanshu Shende completed his Bachelor of Engineering degree in Com-puter Engineering from Modern Education Society’s Wadia College of Engi-neering, Pune.



Rahul Patil completed his Bachelor of Engineering degree in Computer Engineering from Modern Education Society’s Wadia College of Engineering, Pune.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

GrapeNet: An Explainable, Custom Deep Learning Architecture for Early Detection of Grape Leaf Disease

Rashmi Tundalwar[1], Prasenjeet Damodar Patil[2]

Department of Computer Science and Engineering
MIT Art, Design and Technology University, Loni Kalbhor, Pune, India
rtundalwarofficial@gmail.com

Abstract

Speedy, reliable identification of grape leaf diseases is crucial to minimize crop losses and ensure sustainable production. Agronomic manual investigations are sometimes labor-intensive, subjective, and difficult to scale in large-scale vineyards. In this study, we categorize grape leaf diseases using two architectures: GrapeNet, a lightweight CNN designed specifically for grape leaf disease, and a Custom ResNet-50 trained via transfer learning. We use data augmentation throughout to enhance robustness in real-world imaging. The Custom ResNet-50 model achieves an incredibly high test-set accuracy of 99%, whereas GrapeNet achieves 95% accuracy and performs best on most precision, recall, and F1-score metrics. This demonstrates how deep residual networks can be useful for disease detection in precision ag environments for grapevines.

Keywords. Grapevine disease identification, CNN, Deep learning, ResNet50, Transfer learning.

1. INTRODUCTION

Viticulture is a main contributor to global agricultural production, farmers' livelihoods, and sustainable land use. Grapevines are susceptible to several foliar diseases; however, these can delay development, cause poor fruit quality, and affect yields. Early detection of disease is important. Traditional diagnosis methods rely on visual observations by expert agronomists and can be subjective and time-consuming, particularly on a large scale [1], [3]. In recent years, with the development of artificial intelligence and computer vision technology, subtle variations in leaf color have been used to recognize diseases automatically and rapidly. Visual features can be learned from raw images effectively using convolutional neural networks (CNNs). Using pretrained models, such as ResNet, EfficientNet or DenseNet, that have been trained on a large, unbalanced dataset, they can also help in improving the results even when the number of labeled data is small, thereby making real-time monitoring feasible in precision agriculture for scalability and cost-effectiveness [2, 9, 12, 14]. However, difficulties remain, including the risk of overfitting and small validation sets of real-world data [7, 8, 10]. To address these problems, this paper proposes a custom CNN for diagnosing grapevine leaf diseases, GrapeNet. It compares the performance of GrapeNet with that of the transfer-learning-based Custom ResNet-50 and provides an understanding of and validation of the results of both models using explainable AI (XAI).

2. RELATED WORK

Early attempts at computer vision recognition of grapevine diseases used conventional machine learning techniques based on handcrafted features such as color and texture. Classifiers such as Support Vector Machines (SVMs) and Random Forests performed well in controlled settings, but their performance in real-world settings was limited. In recent years, Convolutional Neural Networks (CNNs) have become the preferred way as they learn directly from images. Ji et al. created a Unified Model to classify grape leaf diseases using

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

multiple CNNs [4]. Liu et al. and Math & Dharwadkar proposed deep CNN architectures to detect earlier and with higher accuracy [5] and [7].
New techniques such as capsule networks and attention mechanisms have further improved the reliability of distinguishing similar symptoms by vision. Andrushia et al. integrated a convolutional capsule network (CNN) [6], and Tang et al. combined a lightweight CNN with channel-wise attention to improve efficiency. When labels are limited, transfer learning has been useful using pretrained models such as VGG and ResNet. Noteworthy examples include VGG-based methods for downy and powdery mildew detection [9], improved ResNet-50 using batch normalization [2], and benchmarking different ResNet variants [14]. Some researchers have created synthetic data using Generative Adversarial Networks (GANs) [11] and used deep learning and Particle swarm optimization to classify limited and imbalanced training data [15]. Research applications have been mainly on automatic detection of downy mildew and powdery mildew in the vineyard [13], damage-aware leaf disease classification [3], and grape leaf diagnosis [1]. Despite these developments, data limitations and variability remain, requiring a bespoke CNN-versus-ResNet-50 comparison in addition to explainable AI.

3. METHODOLOGY

This study used 6,189 grape leaf images across six disease categories, split into 75%, 10%, and 15% training, validation, and test sets, respectively, from Mendeley Data. The images were resized, normalized, denoised, and rotated and flipped to improve robustness to variation and better match the models. Two models: a custom CNN and a transfer learning model (Custom ResNet-50) were trained separately, and their outputs were compared using SoftMax.

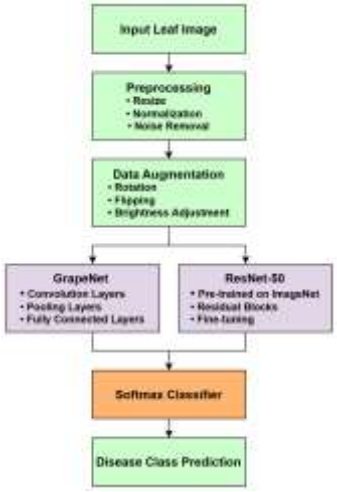
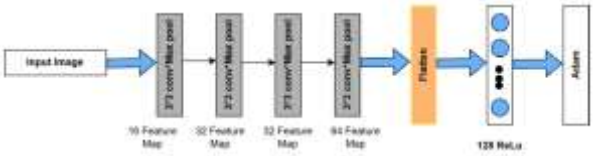


Fig 3.1. Overall architecture of the customized deep-learning-based grapevine leaf disease detection pipeline.

3.1. GrapeNet Custom CNN Architecture

GrapeNet (Fig. 2) comprises four convolutional blocks, each using 3×3 kernels, ReLU activation, and max pooling, with the number of feature maps increasing from 16 to 64 across these blocks, allowing the network to capture progressively more abstract features while remaining lightweight. The resulting feature maps are flattened and passed into a fully connected layer with 128 neurons before the classification head. The whole model is optimized with Adam at a learning rate of 0.001.



All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Fig 3.2. Layer-wise architecture of the GrapeNet custom CNN.

3.2. Custom ResNet-50 Architecture

The Custom ResNet-50 pipeline (Fig. 3) accepts 224×224×3 input images. First, it passes them through an initial 7×7 convolution and max-pooling stage, followed by residual blocks stacked in a 3-4-6-3 arrangement that use a mix of 1×1 and 3×3 convolutions together with skip connections to counter vanishing gradients [2], [9]. Global average pooling then condenses the feature maps before a fully connected classification layer, with Adam used for optimization; together, these improve training stability and overall classification accuracy.

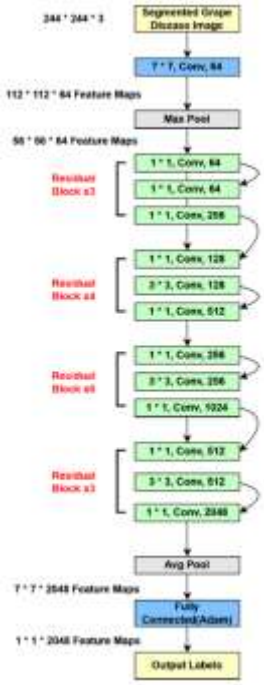


Fig 3.3. Layer-wise architecture of the Custom ResNet-50 model.

3.3. Explainability Using SHAP

To interpret model predictions, we used Shapley Additive Explanations (SHAP) and computed pixel-level attribution scores for both CNN models. The summary plots illustrated the feature importance across the dataset, and the force plots gave instance-by-instance explanations of predictions.

4. RESULTS AND DISCUSSION

GrapeNet and the Custom ResNet-50 were both trained on Google Colab using a Tesla T4 GPU (16 GB) with a batch size of 16 over 10 epochs. To prevent overfitting, we monitored validation performance during training, and we used the held-out test set to assess accuracy, precision, recall, and F1-score after training.

4.1. Confusion Matrix Analysis

Tables II and III share the characteristic of diagonal dominance, indicating that most of the samples were classified correctly for each disease class in both models. GrapeNet has fewer recall points than CPE, but there is little overlap in the classes used for training. GrapeNet has fewer recall points than CPE, but a significant lack of overlap in terms of the classes that are used for training. The Custom ResNet-50, on the other hand, maintains 99% overall accuracy and shows more consistent performance across classes.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Table 1 Performance of grapenet custom cnn model

Class	Precision	Recall	F1-Score	Support
Bacterial Leaf Spot	0.85	0.73	0.79	15
Black Rot	0.96	1.00	0.98	284
Downy Mildew	0.98	0.82	0.89	146
ESCA	1.00	0.97	0.98	288
Healthy Leaves	0.85	0.99	0.91	189
Leaf Blight	1.00	0.99	1.00	259
Powdery Mildew	0.91	0.82	0.86	62
Accuracy		**0.95**		**1243**

Table 2 Performance of custom resnet50 model

Class	Precision	Recall	F1-Score	Support
Bacterial Leaf Spot	0.71	1.00	0.83	15
Black Rot	1.00	1.00	1.00	284
Downy Mildew	0.99	0.98	0.98	146
ESCA	1.00	1.00	1.00	288
Healthy Leaves	1.00	0.98	0.99	189
Leaf Blight	1.00	1.00	1.00	259
Powdery Mildew	0.95	0.95	0.93	62
Accuracy		**0.99**		**1243**

4.2. Significance Analysis

As shown in Figs. 4 & 5, Custom ResNet-50 always outperforms GrapeNet in terms of precision, recall, and F1-score. That 4% increase in accuracy (99% compared to 95%) is significant enough to suggest it is not due to chance. This validates that deeper residual networks with skip connections can be more discriminative for leaf features under the same experimental settings.

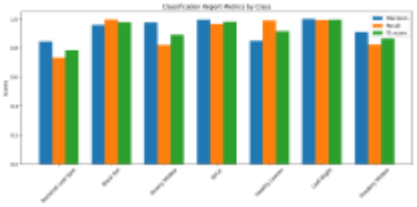


Fig 4. Per-class classification performance of the GrapeNet CNN.

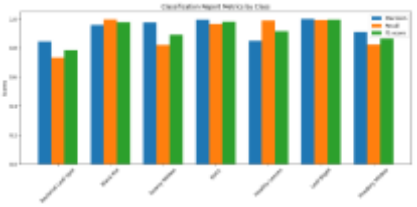


Fig 5. Per-class classification performance of the Custom ResNet-50 model.

5. CONCLUSION

In this work, we have developed a lightweight custom-designed CNN (GrapeNet) and a transfer-learning-based Custom ResNet-50 pipeline to detect grape leaf diseases. GrapeNet was outperformed by the Custom ResNet-50 with a 99% accuracy, due to its deeper residual structure. The precision, recall, and F1-scores were high, though, and GrapeNet was competitive, resilient to a small number of errors between classes that share similar visual features. SHAP-based explainability was used to gain a dataset-wide and case-specific view of the models' predictions. All in all, this pipeline provides a valid, interpretable, scalable, and accurate solution for both precision agriculture and automated monitoring of vineyards and cultivation in the real world.

6. REFERENCES

[1] S. Kunduracioglu and I. Pacal, “Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases,” Journal of Plant Diseases and Protection, vol. 131, no. 3, Jun. 2024.

[2] N. Sagar, K. Suresh, S. Sridhara, B. Patil, C. Archana, Y. S. Sekar, Y. Naveesh, A. Nandan, and R. Sushma, “Precision detection of grapevine downy and powdery mildew diseased leaves and fruits using enhanced ResNet50 with batch normalization,” Computers and Electronics in Agriculture, vol. 232, May 2025.

[3] B. Sai Reddy and S. Neeraja, “Plant leaf disease classification and damage detection system using deep learning models,” Multimedia Tools and Applications, vol. 81, no. 17, Jul. 2022.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

[4] M. Ji, L. Zhang, and Q. Wu, “Automatic grape leaf disease identification via UnitedModel based on multiple convolutional neural networks,” Information Processing in Agriculture, vol. 7, no. 3, 2020.

[5] B. Liu, Z. Ding, L. Tian, D. He, S. Li, and H. Wang, “Grape Leaf Disease Identification Using Improved Deep Convolutional Neural Networks,” Frontiers in Plant Science, vol. 11, Jul. 2020.

[6] Diana Andrushia, T. Mary Neebha, A. Trephena Patricia, S. Umadevi, N. Anand, and A. Varshney, “Image-based disease classification in grape leaves using convolutional capsule network,” Soft Computing, vol. 27, no. 3, pp. 1457–1470, Feb. 2023.

[7] R. M. Math and N. V. Dharwadkar, “Early detection and identification of grape diseases using convolutional neural networks,” Journal of Plant Diseases and Protection, vol. 129, no. 3, pp. 521–532, Jun. 2022.

[8] Z. Tang, J. Yang, Z. Li, and F. Qi, “Grape disease image classification based on lightweight convolution neural networks and channel-wise attention,” Computers and Electronics in Agriculture, vol. 178, Nov. 2020.

[9] Y. Ashurov, M. S. A. M. Al-Gaashani, N. A. Samee, R. Alkanhel, G. Atteia, H. A. Abdallah, and M. Saleh Ali Muthanna, “Enhancing plant disease detection through deep learning: a Depth wise CNN with squeeze and excitation integration and residual skip connections,” Frontiers in Plant Science, vol. 15, Jan. 2025.

[10] Y. Chen and Q. Wu, “Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks,” Precision Agriculture, vol. 24, no. 1, pp. 235–253, Feb. 2023.

[11] S. Mousavi and G. Farahani, “A Novel Enhanced VGG16 Model to Tackle Grapevine Leaves Diseases with Automatic Method,” IEEE Access, vol. 10, pp. 111564–111578, 2022.

[12] L. Ghiani, S. Serra, A. Sassu, A. Deidda, and F. Gambella, “Automated detection of downy mildew and powdery mildew symptoms for vineyard disease management,” Smart Agricultural Technology, vol. 11, p. 100877, Aug. 2025.

[13] S. Vats, J. Anand, V. Kukreja, and R. Sharma, “Deep Learning-based VGG16, VGG19, and ResNet Models for Grapevine Disease Classification,” in 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). Pune, India: IEEE, Apr. 2024.

[14] S. Alsubai, A. K. Dutta, A. H. Alkhayyat, M. M. Jaber, A. H. Abbas, and A. Kumar, “Hybrid deep learning with improved Salp swarm optimization based multi-class grape disease classification model,” Computers and Electrical Engineering, vol. 108, p. 108733, May 2023.

Biographies



Rashmi Ramesh Tundalwar is an Assistant Professor with 14+ years’ teaching experience, specializing in AI, ML, Data Science, and IoT.



Prasenjeet Damodar Patil is a Professor with 16+ years of teaching and research experience, specializing in AI, Machine Learning, IoT, Image Processing, and Data Science.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Ambiguity-Aware Deepfake Image Detection Using StyleGAN2-ADA and CNN with Memory-Guided Recall

Ashrita Joshi[1], Satish Bhadane[2], Raviraj Bhakare [3],Aakanksha Bodke[4], Chaitali Bonde[5], Neha Chaudhari[6]

[1]*Asst. Prof., Computer Engineering, PVGCOE & SSDIOM, Nashik, India, ashrita.joshi@pvgcoenashik.org*
[2]*Asst. Prof., Information Technology, PVGCOE & SSDIOM, Nashik, India, satish.bhadane@pvgcoenashik.org*
[3]*Computer Engineering, PVGCOE & SSDIOM, Nashik, India, ravirajbhakare73@gmail.com*
[4]*Computer Engineering, PVGCOE & SSDIOM, Nashik, India, aakankshabodke19@gmail.com*
[5]*Computer Engineering, PVGCOE & SSDIOM, Nashik, India, chaitalibonde1007@gmail.com*
[6]*Computer Engineering, PVGCOE & SSDIOM, Nashik, India, cneha2404@gmail.com*

Abstract.

With deepfakes being generated using GANs, highly realistic fake images have been created, making it difficult to distinguish them from real images. The use of StyleGAN2-ADA further minimizes visual differences, increasing the complexity of detection. Although convolutional neural networks (CNNs) are widely used for deepfake detection, their performance degrades when differences between real and fake images are minimal. In this paper, we propose an ambiguity-aware deepfake detection model that integrates CNN-based classification with a memory-guided recall mechanism. When prediction confidence is low, previously learned feature representations are recalled to support decision-making. The proposed approach focuses on improving reliability and consistency rather than only maximizing accuracy. This method is particularly useful for real-world applications such as media verification, digital forensics, and cybersecurity analysis.

Keywords. Deepfake Detection; StyleGAN2-ADA; Convolutional Neural Networks; Memory-Guided Recall; Ambiguity-Aware Classification

1. INTRODUCTION

At present, deepfakes have advanced to a level where it is possible to generate synthetic images that appear highly similar to actual visual content. GAN-based models like StyleGAN2-ADA generate very realistic images, making their detection increasingly difficult. In applications requiring authentic images, there is the need for better ways to detect deepfakes. Past literature has shown the fast development in deepfake generation and also the current limitations of detectors in detecting difficult deepfake instances [4], [8]. CNNs are commonly employed for deepfake detection since they learn image features efficiently [1]. The CNN performs well when clear distinctions exist between real and deepfake images. However, as technologies in

image generation advance, it becomes increasingly hard to distinguish between real and fake images. Even though recent studies have examined advancements and improvements in models and methodologies, it has proved difficult to overcome such problems [2], [3].With the aim of resolving this problem, an approach for detecting deepfakes has been presented based on uncertainty in predictions. The chance of misclassification is therefore minimized and reliability is achieved. The technique can also prove to be helpful when trying to tackle complicated borderline cases, whereby the distinction between the two classes becomes extremely minimal.

2. BACKGROUND AND RELATED WORK

Many Convolutional Neural Network (CNN)-based models have been developed because of the ability of CNNs to learn discriminative visual features automatically. These models perform effectively when clear differences exist between real and deepfake images. CNN-based detection methods have shown strong performance in identifying manipulated content by learning spatial and texture-based features from images [1]. However, their effectiveness depends largely on the presence of distinguishable features between real and fake samples. Several studies have attempted to improve deepfake detection capability using hybrid deep learning models involving generative models and advanced feature extraction techniques [2], [3], [6]. Some approaches make use of GAN-based features, frequency domain analysis, and structural inconsistency detection to improve classification performance. Although these methods improve accuracy, they still face challenges when dealing with images that have only minor visual differences.

3. METHODOLOGY

3.1. Overview of the Proposed Framework

The model under consideration involves the use of a CNN for detecting deepfake images, as well as a memory module to overcome the cases of uncertainty in the prediction. The entire process works in stages starting from the preprocessing stage, followed by the feature extraction and classification stages, and ending with dealing with uncertainties in the predictions. The overall workflow is shown in Figure 1.

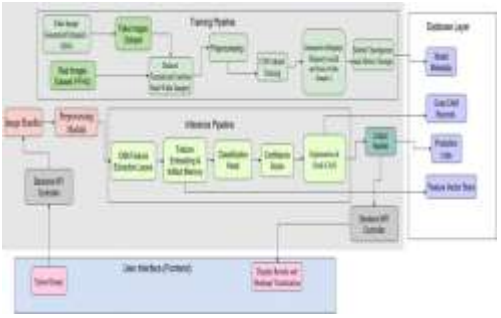


Figure 2.1.Proposed deepfake detection system architecture

3.2. Dataset Description

The dataset utilized in the present experiment comprises real and synthesized faces. Real images are sampled from the FFHQcollection; their number is about 50,000

images of 256x256px size. The generated images are produced by StyleGAN2-ADA, yielding another collection of 50,000 fake images. The total number of images is close to 100,000. The data is split in an 80:20 proportion for training and validation purposes. When training, data goes through random sampling in batch size; thus, during each epoch, the model will not encounter all of the images. As a result, after a few iterations, the model will have seen an estimated total of 40,000 to 50,000 images. There are around 20,000 images that can be found in the validation set, which serves as a means for evaluating the model's performance.

3.3. Data Preprocessing

Each image is resized into 96x96 pixels prior to training. Pixel intensity is normalized using the [0,1] range; in other words, the images are divided by 255. In order to enhance generalization, some basic augmentations are performed, such as horizontal flips and slight rotation.

3.4. CNN-Based Feature Extraction

Convolutional Neural Network is used for extracting features from the images and classifying them into two classes. There are three convolutional layers with filters 32, 64, and 128 respectively, which are followed by max pooling. The feature maps are then flattened and fed to a dense layer with 128 nodes. Afterward, a dropout layer with a probability of 0.5 is applied to avoid overfitting. The final layer uses the sigmoid activation function.

3.5. Ambiguity-Aware Prediction

The output probability of this model is p [0,1]. The predicted results are treated differently based on the confidence level. The results with high confidence are used directly, while those near the decision boundary are viewed as ambiguous and sent to the memory for refining.

3.6. Memory-Guided Recall Mechanism

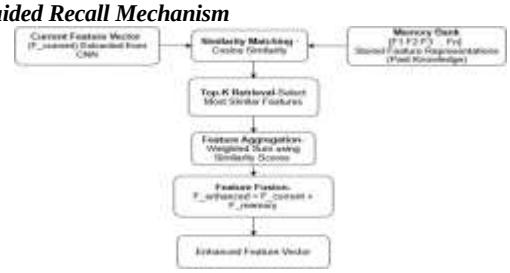


Figure 3.6.2.Memory-Guided Recall Mechanism for Feature Enhancement

4. EXPERIMENT RESULTS

4.1. Experimental Setup

The implementation of the proposed model was carried out using the TensorFlow/Keras framework, while the training and validation were performed using a dataset split of 80% and 20%, respectively. The training process was carried out with a batch size of 32 samples.Additionally, the trained model was tested on unseen real-world images, including AI-generated images, to evaluate its performance under practical conditions.

4.2. Performance Metrics

As demonstrated in Table 1, the proposed model shows an average accuracy of around 78% which indicates a stable and generalized model.

Table 1: Classification Report

Class	Precision	Recall	F1-Score	Support
Real (0)	0.87	0.67	0.76	5,095
Fake (1)	0.73	0.90	0.81	5,145
Accuracy	—	—	**0.78**	**10,240**

4.3. Results and Analysis

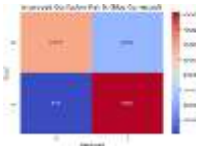


Figure 4.3.1 Confusion matrix of CNN



Figure 4.3.2 Grad-CAM++ Visualization

5. CONCLUSION

Ambiguity arises when there are small variations between real and fake images. In this regard, the current study is concerned about the problem, where an innovative approach of employing a CNN model trained from scratch by utilizing examples generated from StyleGAN2-ADA, in addition to employing a memory recall approach, is used. Results obtained show that the proposed method demonstrates stable performance in solving the assigned task despite the presence of ambiguity, where similarity of features improves decision making.

6. REFERENCES

[1] G. Aggarwal, A. K. Srivastava, K. Jhajharia, N. V. Sharma, and G. Singh, "Detection of Deep Fake Images Using Convolutional Neural Networks," in Proc. 3rd Int. Conf. Technological Advancements in Computational Sciences (ICTACS), IEEE, Nov. 2023, pp. 1083–1087.

[2] P. Sharma, M. Kumar, and H. K. Sharma, "GAN-CNN Ensemble: A Robust Deepfake Detection Model of Social Media Images Using Minimized Catastrophic Forgetting and Generative Replay Technique," Procedia Computer Science, vol. 235,pp. 948–960, 2024.

[3] S. Dasgupta, K. Badal, S. Chittam, M. T. Alam, and K. Roy, "Attention-Enhanced CNN for High-Performance Deepfake De-tection: A Multi-Dataset Study," IEEE Access, vol. 13, pp. 101980–101989, Jun. 2025, doi: 10.1109/ACCESS.2025.3578343.

[4] M. Jaiswal and R. Srivastava, "Visual Deepfake Detection: Review of Techniques, Tools, Limitations, and Future Prospects," Multimedia Tools and Applications, Springer, 2024.

[5] J. Kietzmann, I. Lee, and L. W. McCarthy, "Deepfakes: Trick or Treat? A Critical Review of the Literature on Deepfake Detection," IEEE Transactions on Technology and Society, vol. 3, no. 2, pp. 80–92, Jun. 2022.

[6] S. Banerjee, P. Das, and A. Roy, "CED-DCGAN: Channel and Edge-Based Detection of Deep Fake Images Using GAN Fingerprints," Expert Systems with Applications, vol. 235, pp. 122–134, 2024.

[7] D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proc. IEEE Int. Conf. Advanced Video and Signal-Based Surveillance (AVSS), 2022, pp. 1–6.

[8] C. Pu, H. Jiang, and Y. Liu, "A Survey on Deepfake Detection: Current Challenges and Future Trends," Journal of Information Security and Applications, vol. 73, p. 103405, 2023.

[9] T. Dang-Nguyen, T. Giudice, and G. Boato, "Deepfake Detection Based on Discrepancies Between Face and Head Poses,"Multimedia Tools and Applications, vol. 83, pp. 811–832, 2024.

[10] H. Li, X. Yang, and S. Wang, "Exploring Frequency Domain Features for Deepfake Detection," Pattern Recognition Letters, vol. 168, pp. 47–54, 2023.

Biographies



Ashrita H. Joshi is an Assistant Professor at Pune Vidyarthi Griha's College of Engineering & S. S. Dhamankar Institute of Management, Nashik.



Satish N. Bhadane is an Assistant Professor at Pune Vidyarthi Griha's College of Engineering & S. S. Dhamankar Institute of Management, Nashik.



Raviraj Ramesh Bhakare holds a B.E. in Computer Engineering from Pune Vidyarthi Griha's College of Engineering & S. S. Dhamankar Institute of Management, Nashik.



Aakanksha Balasaheb Bodke received the B.E. degree in Computer Engineering from Pune Vidyarthi Griha's College of Engineering & S. S. Dhamankar Institute of Management, Nashik,



Chaitali Dileep Bonde received the B.E. degree in Computer Engineering from Pune Vidyarthi Griha's College of Engineering and Shrikrishna S Dhamankar Institute of Management, Nashik affiliated with Savitribai Phule Pune University, Pune, India, in 2026.



Neha Atul Chaudhari received the B.E. degree in Computer Engineering from Pune Vidyarthi Griha's College of Engineering and Shrikrishna S Dhamankar Institute of Management, Nashik affiliated with Savitribai Phule Pune University, Pune, India, in 2026.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

IoT-Based Underwater Robotic System for Object Detection and Navigation

Sheetal Ingawale[1], Pooja Gore[2], Amol Bhosale[3], Pratibha Chavan[4], Mamata Pawar[5]

[1]Lecturer @Yashoda Technical Campus, Satara, India
[2]Assistant Professor@Karmaveer Bhaurao Patil College Of Engineering, Satara, India
[3]Assistant Professor@Trinity College of Engineering and Research, Pune, India
[4]Professor @Trinity College of Engineering and Research, Pune, India
[5]Lecturer@ Karmaveer Bhaurao Patil Polytechnic, Satara, India

Abstract.

This paper presents the design and testing of a low-cost underwater robot used for monitoring applications. The system is built using an ESP32-CAM module , which allows real-time video streaming and wireless control through a mobile application. A camera captures underwater visuals, and the frames are implemented using basic techniques like grayscale , conversion, edge detection and contour detection to identify objects such as pipes and debris. Along with this , proximity and depth sensors are used to improve navigation and detect obstacles. The robot moves using thrusters controlled through a motor driver circuit. The setup was tested in a small water tank under different conditions. During testing ,it showed stable movement and was able to detect objects with reasonable accuracy. From the testing, it can be seen that the system is suitable for small-scale underwater tasks. The underwater robot is designed with focus on simplicity, cost-effectiveness, and ease of operation , the designed robot can be extended for practical real-world applications.

Keywords. Underwater Robotics, IoT, Object Detection, Autonomous Navigation, Underwater Communication, Computer Vision, Robotic Systems

1. INTRODUCTION

The underwater environment faces many challenges like poor visibility and safety risks. Because of this robots are often used instead of humans for inspection tasks. These robots are commonly used in areas like pipelines , tanks, and underwater structures . Most low-cost systems available today depend on manual control, which makes them harder to operate continuously . Also many of these systems do not include real-time object detection features. This limits their usefulness in practical situations . To minimise these problems our work focuses on developing a simple underwater robotics system that can assist the user by providing basic object detection and easier control.

2. LITERATURE SURVEY

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Recent studies have explored underwater robotic systems for inspection and monitoring tasks. Advanced systems provide high accuracy but require expensive hardware and complex algorithms[1],[3]. Low-cost systems focus on simplicity but often lack real-time detection capability[4]. Image-processing techniques such as edge detection and contour analysis have been used for object detection in underwater environments[2],[6]. Embedded systems like Raspberry Pi and microcontrollers have also been used for cost-effective solutions[4],[7].This research indicates that most systems either focus on high-cost solutions or lack real-time detection capability in low-cost environments. Due to this, we need a simple and affordable robot that detects objects.

3. KEY CONTRIBUTION

The main contributions of this work are as follows:

- Development of a low-cost underwater robotic system using ESP32-CAM
- Integration of image processing and sensor-based navigation for object detection
- Design of a simple and efficient system suitable for small-scale underwater inspection tasks
- Implementation of a real-time monitoring system with mobile-based control

4. EXISTING SYSTEM LIMITATION

Existing underwater robotic systems are mostly dependent on manual operation. This increases operator workload and reduces efficiency. Many systems do not support real-time object detection, making it difficult to identify underwater objects automatically. Communication delay and signal loss are also common issues. Additionally, limited sensor integration reduces the system's ability to adapt to environmental changes.

5. METHODOLOGY



Fig 5.1. Methodology flowchart

The workflow of the system is shown in Fig. 1. The working of the robot is divided into simple steps. The architecture begins by capturing underwater video, which is converted to grayscale and processed using edge and contour detection. After that, edge detection is applied to highlight object boundaries. To identify objects, contour detection is used. At the same time , proximity sensors keep checking for nearby obstacles. If something is detected within a certain distance , the robot either stops or changes its direction.All these operations

are handled by the ESP32-CAM module, which also manages communication with the mobile application.

6. SYSTEM DESIGN

The design of the underwater robot integrates multiple components to improve performance

1. Camera for real-time monitoring
2. Thrusters for movement
3. Sensors for environmental data
4. Mobile interface for control. This combination improves usability and allows better control compared to traditional systems.

7 . OBJECT DETECTION:

A lightweight image-processing approach is used to reduce computational requirements on embedded hardware. Captured video frames are converted into grayscale and processed using edge detection techniques. Contour detection is used to identify object boundaries. This method is suitable for embedded systems and provides faster execution with lower consumption [4], [6]. While less accurate than deep learning methods, it provides sufficient performance for basic detection tasks.

7.1. Navigation

The framework of our system includes semi-autonomous navigation using proximity sensors. When an obstacle is detected, the robot automatically changes direction or alerts the operator. This feature improves safety and reduces manual effort during operation.

8 . System Architecture

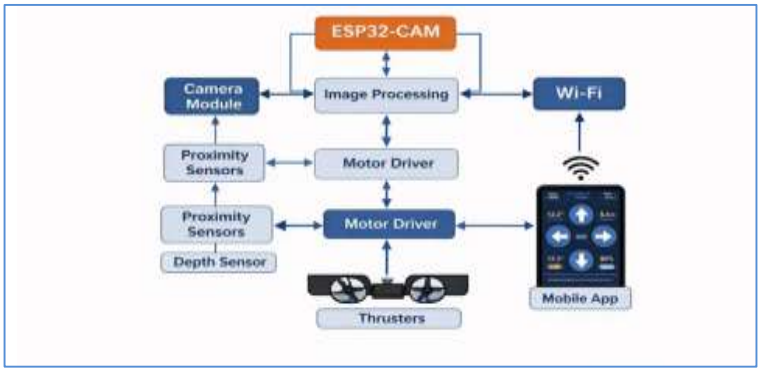


Fig 8.1. System architecture

The system architecture is shown in Fig. 2. The architecture of system of:

- ESP32-CAM microcontroller
- Camera module
- Motor drivers and thrusters
- Sensors

- Mobile control system

All these components help to detect underwater objects.

9. EXPERIMENTAL RESULTS

The experiment was conducted in a controlled water tank environment where different objects such as pipes and solid obstacles were placed to simulate real conditions. The system was tested multiple times under similar conditions to ensure consistency. Detection was considered successful when object boundaries were clearly identified in the processed output. The performance was evaluated based on detection success rate and response time.

Table 1. Experimental results

Parameter	Value
Test Environment	Water tank
Number of Trials	20
Successful Detection	16
Accuracy	80%
Response Time	1.5 sec

10. CONCLUSION

In this work, a low-cost underwater robotic system was developed and tested for basic inspection tasks. The system combines image processing and sensor-based navigation to reduce manual effort and improve usability. During testing in a controlled water tank, the robot showed stable movement and was able to detect objects in most cases. It was also observed that detection accuracy decreases when water clarity is low. The current design is simple and suitable for small-scale applications where cost and ease of use are important. Although the system is basic, it provides a strong foundation for further development. Future improvements can include better detection techniques, improved sensors, and more reliable communication systems.

11. REFERENCES

[1] F. Nauert and P. Kampmann, “Inspection and maintenance of industrial infrastructure with autonomous underwater robots,” *Frontiers in Robotics and AI*, 2023.

[2] Khan, et al., “Subsea pipeline corrosion estimation using image enhancement,” *IEEE Access*, vol. 6, pp. 50609–50621, 2018.

[3] S. Kartal and R. Cantekin, “Autonomous underwater pipe damage detection,” *Journal of Marine Science and Engineering*, 2024.

[4] M. Aliff, et al., “Underwater crack detection using edge detection,” *International Journal of Engineering and Technology*, 2022.
[5] B. Shirose, et al., “Design of ROV for underwater inspection,” in *Proc. IEEE Conference*, 2021.
[6] Y. Al Mashhadany, et al., “Underwater robot with video monitoring,” in *Proc. International Conference*, 2020.
[7] Y. Wang, et al., “Smartphone-based underwater monitoring system,” in *Proc. ACM Conference*, 2015.
[8] T. Shahria, et al., “Underwater rescue robot,” in *Proc. IEEE Conference*, 2019.
[9] G. Martos, et al., “ROV design and control,” B.S. thesis, 2013.
[10] S. Ohata, et al., “AquaBox underwater robot system,” 2007.
[11] Y. Wei, et al., “Mini underwater robot design,” *Applied Mechanics and Materials*, 2013.
[12] R. Govindarajan, et al., “Underwater robot control system,” *International Journal of Science, Engineering and Technology*, 2013.
[13] M. Chitre, et al., “Underwater acoustic communications and networking,” *IEEE Communications Magazine*, 2008.
[14] J. Heidemann, et al., “Research challenges in underwater sensor networks,” *IEEE Wireless Communications*, 2012.
[15] P. Corke, et al., “Experiments with underwater robot localization,” *IEEE Robotics and Automation Letters*, 2016.
[16] R. Sutton, et al., “Autonomous underwater vehicles: Challenges and advances,” *Ocean Engineering*, 2011.

Biographies



Mrs Sheetal J Ingawale She holds a B.E. in Electronics and Telecommunication Engineering and an M.E. in Digital Systems. She is currently an Assistant Professor in the Department of Electronics and Telecommunication Engineering at YTC Satara.



Prof. Pooja Vitthal Gore received the bachelor's degree in Information Technology from SGBAUA University in 2022, the master's degree in computer engineering from DBATU in 2026. She is currently working as an Assistant Professor at the Department of Computer Engineering, Faculty of Engineering, Karmveer Bhaurao Patil College of Engineering Satara .Her research areas include Public security, deep learning, and Computer Vision .

AI-Based Agricultural Robot for Real-Time Crop Monitoring, Disease Detection and Smart Irrigation

Karishma Pimpale [1] Youraj Kanase[2] Sunita Mane[3]

Department of Electronics Engineering,@Karmaveer Bhaurao Patil College of Engineering, Satara,
karishmapimpale311297@gmail.com, yuraj.kanase@kbpcoes.edu.in, sunita.mane@kbpcoes.edu.in

Abstract.

Agriculture requires timely monitoring of crop health, soil condition, and water availability. In many farms, these tasks are still done manually, which may delay disease identification and irrigation decisions. The current paper is about an AI-based agricultural robot that monitors crops, detects plant diseases, and controls irrigation. The study employs IoT sensors, an ESP32 controller, wireless communication, and a CNN-based model to detect plant diseases using leaf images. Irrigation is automatically controlled by soil moisture data, and temperature, humidity, and light detection allow monitoring of the situation on the field. This study will help to minimize manual work, enhance response time, and assist farmers in making decisions faster. Experimental comparison reveals that the time of disease detection, water consumption, labor demand, and disease detection accuracy are improved. It is a low-cost, modular design and is scalable with cloud monitoring and edge AI in the future. % accuracy in disease detection and offers a scalable solution to precision agriculture.

Keywords. Deep Learning, Precision Agriculture, Internet of Things (IoT), Smart Irrigation, Agricultural Robotics.

1. INTRODUCTION

Agriculture is vital to food production, but many tasks require manual observation. Farmers usually inspect crops visually and decide on irrigation based on experience. This method works for small areas, but it becomes difficult when the field size increases or when disease symptoms appear at an early stage. Delay in detecting plant disease reduces crop quality. In the same way, improper irrigation may either waste water or affect plant growth. With the development of IoT, artificial intelligence, embedded systems, and agricultural robotics, smart farming systems can now collect field data and support faster decision-making in farm operations [1], [5], [7]. A small agricultural robot can move in the field, collect sensor readings, capture plant images, and provide alerts when abnormal conditions are found.

Small and medium-sized farmers can benefit from an affordable, easy-to-use system. This paper introduces an AI-based agricultural robot that monitors crops, detects diseases, and intelligently irrigates. The system includes an ESP32 controller, soil moisture sensor, temperature and humidity sensor, light sensor, ultrasonic sensor, and a CNN disease classification model. The overall objective of the system is to minimize manual monitoring and deliver up-to-date data on crop and soil conditions.

2. LITERATURE SURVEY

Several researchers have worked on agricultural robotics, IoT-based irrigation, and plant disease detection. Agricultural robots are mainly used to reduce repetitive field work, support field monitoring, and improve decision-making in precision agriculture [1], [2]. Earlier studies also show that mobile robots can help in collecting field data and assisting crop inspection activities. Deep learning can detect plant diseases from leaf images using CNN-based models trained on proper datasets [3], [4].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Disease classification using images performs well when the images used are clear, well-labeled and when taken in controlled conditions . Nevertheless, there are real field conditions that can interfere with performance due to lighting conditions, background noise, camera angle, and leaf position.

The irrigation systems based on IoT are common in water management. Such systems monitor the soil moisture and weather and then turn on irrigation, thus preventing unnecessary water consumption as opposed to fixed-time irrigation [5], [8]. Many current systems are only concerned with a single task, e.g. agro irrigation or detecting plant diseases. The suggested project integrates crop detection, intelligent irrigation, and disease identification into one inexpensive robot

3. KEY CONTRIBUTION

The miniature agricultural robot with sensing, disease detection, and irrigation control is the main contribution. This study measures field conditions with soil moisture, temperature, humidity, light, and ultrasonic sensors. An ESP32 controller processes sensor data and controls the irrigation pump. Leaf images are used to detect plant disease using a CNN model. The system also allows communication wirelessly to allow readings and alerts to be sent to a dashboard. Design is modular for future updates.

4. MACHINE LEARNING MODEL

A CNN-based deep learning model is used for disease detection.

- Architecture: VGG16-based CNN
- Dataset: PlantVillage (18,000+ images)
- Classes: 9 diseases + healthy
- Accuracy: ~90%

The model processes real-time images and detects diseases within seconds

5. METHODOLOGY

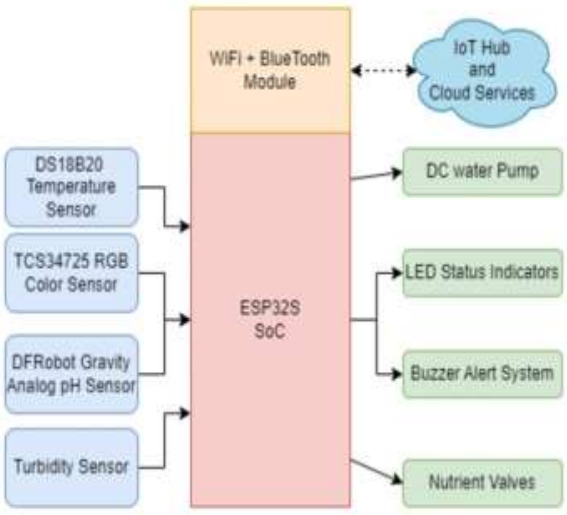


Fig 5.1. Methodology Flowchart

The system operates as follows:

1. Collect sensor data
2. Capture plant image
3. Preprocess image
4. Apply CNN model
5. Detect disease
6. Execute actions (irrigation/alerts)

Algorithm 1: Smart Irrigation Control

Input: Soil Moisture (M)
If M < Threshold → Turn ON Pump.
Else → Turn OFF Pump

6. SYSTEM DESIGN

The system is based on a multi-layer architecture comprising sensor, processing, communication, and application layers.

6.1 ***Sensor Layer***

This layer collects environmental information with:

- Soil moisture sensor
- Humidity and temperature sensor (DHT22)
- Light sensor (BH1750)
- Ultrasonic sensor

6.2 Processing Layer

The sensor information is processed on the ESP32 microcontroller, and decision-making is done. It regulates the irrigation and the movement of the robots according to the set values.

6.3 Communication Layer

Wi-Fi is used to facilitate wireless communication. MQTT/HTTP is used to transfer data to cloud platforms.

6*.4 Application Layer*

The application allows farmers to keep track of conditions and alerts.

7. RESULTS AND PERFORMANCE ANALYSIS

The system was evaluated against traditional farming methods.

Table 1 Experimental results

Parameter	Traditional	Proposed
Detection Time	1.0	0.01
Water Usage	1.0	0.65
Labor Cost	1.0	0.10
Accuracy	0.65	0.90

7.1. Key Improvements

- Disease detection time reduced from days to seconds
- Water usage reduced by 30–40%
- Labor cost reduced by up to 90%
- Detection accuracy improved to ~90%

These results demonstrate improved efficiency and reliability of the system.

8. DISCUSSION

AI, IoT, and robotics enhance the functionality of the overall agricultural activities. The system helps to make decisions faster and minimizes the use of manual labor. They will be subjected to frequent surveillance to offer timely intervention, thereby minimizing the losses that crops will suffer. The performance of the system can, however, change according to the environmental factors, stability of the network and model accuracy. Further innovations can add to resiliency and flexibility.

9. CONCLUSION

The article presents a concept of an artificial intelligence-based agricultural robot that would be used to improve crop monitoring, detection of diseases, and irrigation. The system demonstrates that it is significantly quicker, more precise and better resource utilization than those of the past. The solution combines sensing and intelligence with automation to provide a viable solution to modern precision agriculture. It has strong potential for real-world adoption, especially in resource-constrained farming environments.

10. REFERENCES

[1] R. Shamshiri, C. Weltzien, I. A. Hameed, I. J. Yule, T. E. Grift, S. K. Balasundram, L. Pitonakova, D. Ahmad, and G. Chowdhary, "Research and development in agricultural robotics: A perspective of digital farming," International Journal of Agricultural and Biological Engineering, vol. 11, no. 4, pp. 1–14, 2018.

[2] D. F. Yepez-Ponce, J. M. Garcia-Lara, and A. R. Mendoza, "Mobile robotics in smart farming: Current trends and applications," Agronomy, vol. 13, no. 9, pp. 1–20, 2023.

[3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, pp. 1–10, 2016.

[4] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[5] K. Sharma, M. Sharma, and A. Gupta, "Integration of artificial intelligence and IoT in precision agriculture," Smart Agricultural Technology, vol. 5, pp. 100–115, 2024.

[6] A. Ahmad, S. Saraswat, and R. K. Gupta, "MobileNet-based plant disease detection for smart agriculture," International Journal of Computer Applications, vol. 182, no. 45, pp. 1–6, 2019.

[7] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[8] N. Kumar, V. Kumar, and S. Patel, "IoT-based smart irrigation system for efficient water management," Sensors, vol. 20, no. 12, pp. 1–15, 2020

[9] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," Computer Networks, vol. 38, no. 4, pp. 393–422, 2002.

[10] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard computing: Sensor networks in agricultural production," IEEE Pervasive Computing, vol. 3, no. 1, pp. 38–45, 2004.

[11] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[12] C. Zhang, N. Wang, and M. Wang, "Precision agriculture—A worldwide overview," Computers and Electronics in Agriculture, vol. 36, no. 2–3, pp. 113–132, 2002.

[13] A. Pretto, E. Pagello, and E. Menegatti, "Aerial-ground robotic system for precision farming," arXiv preprint arXiv:1911.03098, 2019.

[14] L. Droukas, D. Kateris, and D. Bochtis, "A survey of robotic harvesting systems and enabling technologies," arXiv preprint arXiv:2207.10457, 2022.

[15] M. R. Islam, M. R. Islam, and M. R. Rahman, "Smart farming automation system using IoT," International Journal of Engineering Research & Technology, vol. 9, no. 5, pp. 1–5, 2020.

[16] S. R. Gengaje and S. M. Deshmukh, "Arduino-based agricultural robot for smart farming," International Journal of Innovative Research in Science, Engineering and Technology, vol. 6, no. 5, pp. 1–6, 2017.

Biographies



Ms. Karishma L. Pimpale received her bachelor's degree in Electronics and Telecommunication (E&TC) Engineering from Shivaji University in 2020. She is currently pursuing her Master of Technology (M.Tech.) degree at Karmaveer Bhaurao Patil College of Engineering satara, affiliated with Dr. Babasaheb Ambedkar Technological University (DBATU), with a specialization in Electronics. Her areas of interest include electronics, research, teaching, and emerging technologies.



Dr.Yuvraj K. Kanse received his bachelor's degree in Electronics Engineering from Shivaji University , his master's degree in Electronics engineering from Shivaji University and his PhD degree in Electronics engineering from Shivaji University , respectively. He is currently working as an Assistant Professor in the Department of Electronics and Telecommunication Engineering at Karmaveer Bhaurao Patil College of Engineering satara. His area of expertise communication, power electronics, medical electronics



Dr.Sunita V. Mane received her bachelor's degree in Electronics Engineering from Shivaji University , her master's degree in Electronics engineering from Shivaji University and her PhD degree in Electronics engineering from Shivaji University , respectively. She is currently working as an Assistant Professor in the Department of Electronics and Telecommunication Engineering at Karmaveer Bhaurao Patil College of Engineering satara. Her area of expertise antenna design

Wireless AI-Based Biomedical Parameter Monitoring System for Human Health and Fatigue Detection

Priyanka Pawar[1] Sunita Mane[2] Supriya Kadam[3]

Department of Electronics Engineering,@Karmaveer Bhaurao Patil College of Engineering, Satara,
pawarpriyanka677@gmail.com,sunita.mane@kbpcoes.edu.in,
supriya.kadam@kbpcoes.edu.in

Abstract.

Continuous monitoring of physiological parameters is essential for improving healthcare outcomes and ensuring occupational safety in high-risk environments. Traditional monitoring methods rely on periodic clinical assessments, which are insufficient for detecting real-time physiological changes. This paper presents a wireless AI-based biomedical monitoring system that integrates wearable sensors, embedded processing, and intelligent analytics to enable real-time health monitoring and fatigue detection. The system collects physiological data such as heart rate, oxygen saturation, temperature, and stress indicators using wearable devices. These signals are processed through an embedded controller and analysed using lightweight machine learning techniques. A vision-based module enhances fatigue detection by identifying behavioural indicators such as eye closure and yawning. The processed data is transmitted to a cloud-based dashboard, enabling remote monitoring and alert generation.
The results indicate that the proposed system achieves improved accuracy, reduced latency, and efficient real-time performance, making it suitable for industrial safety and remote healthcare applications.

Keywords. Biomedical Monitoring, Internet of Things (IoT), Wearable Sensors, Edge AI, Fatigue Detection, ESP32, Wireless Communication.

1. INTRODUCTION

Industrial environments require continuous attention, physical effort, and quick decision-making. Fatigue and physiological imbalance significantly affect human performance and increase the likelihood of accidents. Traditional monitoring approaches rely on periodic medical evaluation, which fails to capture real-time physiological variations. Wearable sensor technology has enabled continuous monitoring of vital parameters such as heart rate and stress levels. When combined with Artificial Intelligence (AI), these systems can identify abnormal patterns and provide early warnings.

Edge computing further enhances system performance by processing data locally, reducing latency and preserving privacy [1]. Physiological indicators such as heart rate variability (HRV) and galvanic skin response (GSR) have been widely used for fatigue detection [3]. Integrating these signals with behavioral indicators such as eye closure improves detection accuracy.

This paper proposes a hybrid system combining wearable sensing, computer vision, and AI-based processing to enable real-time fatigue detection and health monitoring.

2. LITERATURE SURVEY

Modern wearable devices enable continuous monitoring of physiological signals, allowing early detection of abnormal health conditions and improving diagnostic efficiency [1].

These systems typically consist of sensing, processing, and communication layers that support real-time health tracking. Energy-efficient embedded systems are essential for wearable applications. Lee and Al Faruque demonstrated optimized processing techniques for low-power biomedical devices [2].

Machine learning models further enhance system intelligence by analyzing complex physiological data patterns [3].

Edge AI has gained attention for real-time applications, allowing models to run directly on embedded systems [4]. Vision-based fatigue detection using facial features has also shown promising results [7]. Despite advancements, challenges such as data privacy, sensor accuracy, and energy efficiency remain critical.

3. KEY CONTRIBUTION

This paper makes the following key contributions to the field of AI-based healthcare monitoring and fatigue detection:

1. Integrated Multimodal Monitoring Framework: A unified system is developed that combines physiological sensing (HR, SpO_2, temperature, GSR) with vision-based fatigue detection, enabling more accurate and reliable assessment of human health conditions compared to single-modality systems.
2. Edge AI–Enabled Real-Time Processing: The proposed system deploys lightweight machine learning models on an embedded platform (ESP32), enabling real-time decision-making with reduced latency and improved data privacy, minimizing reliance on cloud infrastructure.
3. Hybrid Architecture for Scalable Monitoring: A layered architecture integrating edge computing and cloud analytics is designed to support both real-time local processing and long-term data analysis, ensuring scalability and efficient resource utilization.
4. Multi-Level Intelligent Alert Mechanism: The system introduces a hierarchical alert system (local alerts, mobile notifications, and dashboard warnings) for early detection of fatigue and stress, improving safety in industrial and healthcare environments.
5. Efficient Data Processing Pipeline: A complete workflow including data acquisition, pre-processing, feature extraction, and classification is implemented to ensure reliable analysis of both physiological and behavioural signals.
6. Low-Cost and Practical Implementation the use of affordable sensors and embedded hardware makes the system cost-effective and suitable for real-world deployment, especially in resource-constrained environments.

4. SYSTEM ARCHITECTURE

The system follows a layered architecture consisting of:

- Sensor Layer: Collects physiological data (HR, SpO_2, temperature, GSR)
- Edge Layer: ESP32 performs pre-processing and feature extraction
- AI Layer: Machine learning model classifies user state
- Communication Layer: Wi-Fi transmits data
- Cloud Layer: Stores and analyses data
- Application Layer: Dashboard displays results

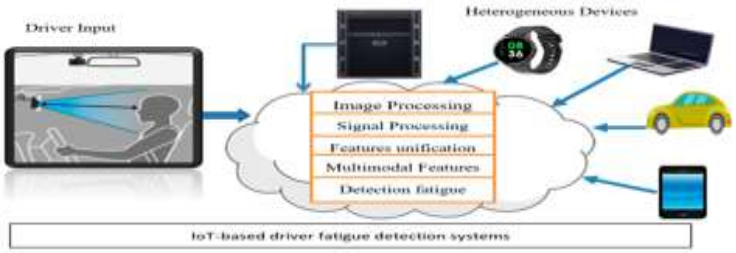


Fig 4.1. System Architecture of the proposed AI-based biomedical monitoring system.

5. METHODOLOGY

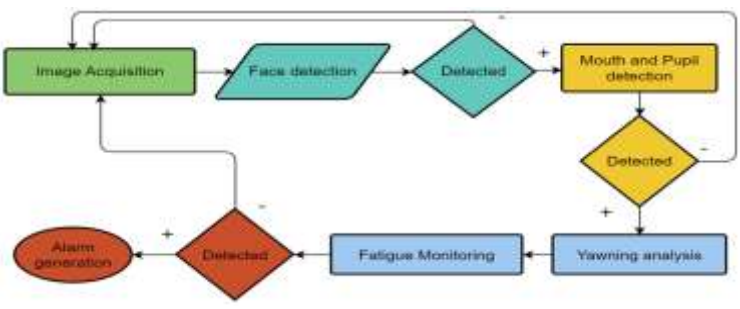


Fig 5.2. Methodology workflow for fatigue detection and health monitoring.

The system operates through the following steps:

1. Data Acquisition: Sensors collect physiological signals and camera captures facial data
2. Pre-processing: Noise removal and normalization
3. Feature Extraction: HRV, stress level, blink rate, yawning frequency
4. AI Classification: Model classifies user condition
5. Communication: Data transmitted using Wi-Fi/MQTT
6. Visualization: Dashboard displays real-time results
7. Alert Generation: Alerts triggered based on fatigue level

6. PROPOSED SYSTEM

- MAX30102 Sensor: Measures heart rate and oxygen saturation (SpO_2), providing key indicators of cardiovascular health.
- GSR Sensor: Monitors skin conductance to estimate stress and emotional response levels.
- Temperature Sensor: Tracks body temperature variations, helping detect abnormal physical conditions.
- ESP32 Controller: Acts as the main processing unit, handling data acquisition, pre-processing, and execution of the AI model.
- Camera Module: Captures facial images to analyse fatigue-related features such as eye closure and yawning.
- Edge AI Model: Performs real-time classification of user conditions (normal, fatigued, stressed) directly on the device.
- Cloud Database: Stores collected data and supports remote monitoring and long-term analysis.
- Dashboard Interface: Displays real-time health parameters, fatigue levels, and alerts for users and supervisors.

7. RESULTS AND PERFORMANCE ANALYSIS

The system was evaluated using collected physiological and visual data under controlled conditions.

Table 1 Experimental results

Parameter	Value
Accuracy	95.2%
Precision	92.8%
Recall	94.1%
F1 Score	93.4%
Response Time	1.3 sec

The results show that the system achieves high accuracy in identifying fatigue conditions, demonstrating reliable classification of physiological and behavioural states. The integration of multiple data sources improved performance compared to single-modality systems. The response time confirms the efficiency of edge processing.

8. DISCUSSION

The experimental results indicate that the proposed system effectively detects fatigue and monitors physiological conditions in real time. The use of both physiological signals and vision-based features improves detection accuracy compared to single-modality approaches [3].

Edge-based processing on the ESP32 reduces latency and enables quick response, making the system suitable for real-time applications [1]. The wireless communication module also demonstrates reliable data transmission with minimal delay, supporting continuous monitoring through the cloud platform [7].

However, minor limitations were observed due to sensor noise and variations in lighting conditions affecting the vision module. Despite these challenges, the system provides a balanced performance in terms of accuracy, speed, and efficiency.

9. CONCLUSION

The proposed AI-based biomedical monitoring system provides an efficient solution for real-time health monitoring and fatigue detection. By integrating wearable sensors, edge computing, and AI analytics, the system enables proactive healthcare management and improved occupational safety.

10. REFERENCES

[1] Hefnawy, M. El-Bendary, and A. Hassan,
"Wearable IoT-based healthcare monitoring systems: Architecture and applications,"
IEEE Internet of Things Journal, vol. 10, no. 5, pp. 4123–4135, 2023.

[2] S. Lee and M. A. Al Faruque, "Energy-efficient biomedical monitoring using embedded wearable platforms,"
IEEE Transactions on Biomedical Circuits and Systems, vol. 16, no. 4, pp. 789–802, 2022.

[3] K. P. Reddy, R. Kumar, and S. Sharma,
"Machine learning enabled physiological monitoring using wearable sensors,"
IEEE Sensors Journal, vol. 23, no. 18, pp. 20541–20552, 2023.

[4] T. Zhang, Y. Liu, and J. Wang,"Lightweight machine learning frameworks for edge-based biomedical monitoring,"
in Proc. Machine Learning and Systems (MLSys), 2022, pp. 115–126.

[5] World Health Organization (WHO),
Global Digital Health and Remote Monitoring Report, Geneva, Switzerland, 2023.

[6] National Institute for Occupational Safety and Health (NIOSH),
Workplace Fatigue and Health Monitoring Systems Report, 2024.

[7] J. Wang, L. Chen, and H. Zhao,
"AI-enabled Internet of Medical Things framework for wearable health monitoring,"
IEEE Access, vol. 11, pp. 65432–65445, 2023.

[8] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. S. Kwak,
"Internet of Things for health care: A comprehensive survey,"
IEEE Access, vol. 3, pp. 678–708, 2015.

[9] A. Pantelopoulos and N. G. Bourbakis,"A survey on wearable sensor-based systems for health monitoring and prognosis,"
IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 1, pp. 1–12, 2010.

[10] M. Chen, Y. Ma, J. Song, C. F. Lai, and B. Hu,
"Smart clothing: Connecting human with clouds and big data for sustainable health monitoring,"
IEEE Network, vol. 30, no. 5, pp. 38–43, 2016.

[11] Z. Zhang, J. Liu, and X. Chen,"Driver fatigue detection based on eye state analysis using machine learning,"
IEEE Access, vol. 7, pp. 118727–118738, 2019.

[12] X. Wang, Y. Han, V. Leung, D. Niyato, X. Yan, and X. Chen,
"Convergence of edge computing and deep learning: A comprehensive survey,"
IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 869–904, 2020.

Biographies



Ms. Priyanka S. Pawar received her bachelor's degree in Electronics Engineering from Shivaji University in 2014. She is currently pursuing her Master of Technology (M.Tech.) degree at Karmaveer Bhaurao Patil College of Engineering satara, affiliated with Dr. Babasaheb Ambedkar Technological University (DBATU), with a specialization in Electronics. Her areas of interest include electronics, VLSI, power electronics, and emerging technologies.



Dr.Sunita V. Mane received her bachelor's degree in Electronics Engineering from Shivaji University , her master's degree in Electronics engineering from Shivaji University and her PhD degree in Electronics engineering from Shivaji University , respectively. She is currently working as an Assistant Professor in the Department of Electronics and Telecommunication Engineering at Karmaveer Bhaurao Patil College of Engineering satara. Her area of expertise antenna design



Dr.Supriya P. Salave received her bachelor's degree in Electronics and Telecommunication Engineering from Shivaji University , her master's degree in Electronics engineering from Shivaji University and her PhD degree in Electronics engineering from Shivaji University , respectively. She is currently working as an Assistant Professor in the Department of Electronics and Telecommunication Engineering at Karmaveer Bhaurao Patil College of Engineering satara. Her area of expertise embedded system, electromagnetic enggineering, medical electronics.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

AgriSense: An AI-Based Smart Agriculture Web Platform for Crop Disease Analysis and Decision Support

Sagar Rajebhosale[1], Suraj Ghule[2], Rohan Gadhave[3], Avneet Kaur[4], Swati Paneri[5]

[1,2,3,4,5]*Keystone School of Engineering, Pune, India, ssrajebhosale@gmail.com, surajghule8037@gmail.com, rohangadhave02@gmail.com, avneetbakshi152@gmail.com, swatijoshipaneri@email.com*

Abstract.

Farming has always been faces challenges, and things like sudden weather changes and quickly spreading crop diseases just make it worse and cause farmers to lose. These problems directly affect how much farmers are able to produce. This paper talks about AgriSense, a web platform that helps in detecting crop diseases and supports better farming decisions.At the center ,the system uses a deep learning model that looks at images of plant leaves to find diseases. It uses IOT Data instead of real sensors and hardware. The platform is built using Java and is kept simple, so farmers can check crop conditions easily and get useful suggestions which makes our user friendly.During the testing, the model performed well in to identifying diseases under controlled conditions. Along with detection, AgriSense also provides suggestions for irrigation and pesticide use, which helps in reducing waste and time and using resources more effectively.This platform is made for smart farming where cost is low and scaling up matters. In the future, the real-time IoT hardware can be added to improve automation.

Keywords. Artificial Intelligence; Smart Agriculture; Crop Disease Detection; Machine Learning; Web Application; Precision Farming.

1. INTRODUCTION

Farming is a vital sector of the Indian economy, with a large population depending on agriculture. However, farmers face challenges such as changing weather, resource wastage, and unexpected crop diseases. AI-based image analysis can help detect crop diseases, while IoT technologies monitor temperature, soil moisture, and humidity [1], [2], [3]. However, many existing solutions require costly hardware and focus on specific problems. AgriSense addresses this gap through a simple web-based platform that allows farmers to upload crop images, receive disease analysis, and get practical recommendations for irrigation and crop care without expensive hardware. The platform, developed using Java, provides a user-friendly interface requiring minimal technical knowledge. Although it is currently a working prototype, its design allows future integration of IoT devices without major system changes.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. LITERATURE SURVEY

Quite a few researchers have looked into using AI and IoT for smarter agriculture. Some of that work [1] and [2], for example, deals with automating the irrigation and spotting crop diseases using machine learning. Others, like [3] and [4], bring in cloud systems and physical sensors to track field conditions and help farmers to decide what to do next. Good work, but a lot of these setups need serious hardware on the ground, which makes it costly and complex to actually run. Later, things are shifting now. More research is now going towards web-based platforms: nothing fancy, no heavy equipment, just something that opens in a browser and works.

TABLE I: Literature Survey on Smart Agriculture Systems

No.	Title	Key Contribution
5	Low Power IoT Networks for Agricultural Monitoring Using LoRaWAN	Uses LoRaWAN protocol for long-range, low-power communication suitable for large-scale and remote farms.
6	Hybrid AI Models for Crop Disease Classification	Combines CNN and SVM to improve classification accuracy and processing speed over traditional models.
7	Java-Based Middleware for IoT Data Management in Agriculture	Middleware platform enhancing IoT data interoperability, scalability, and real-time analytics for decision support.
8	IoT-based Crop Monitoring and Disease Detection	Low-cost IoT system with sensors and ESP32-CAM for monitoring crops and performing ML-based disease detection.
9	AI and IoT-powered Edge Device for Crop Pest and Disease Detection	Portable edge device using lightweight CNN (Tiny-LiteNet) for real-time disease detection with high accuracy.
10	AI and IoT for Plant Disease Identification (Survey)	Survey of AI and IoT approaches including ML, CNN, datasets, and challenges in real- world deployment.

3. PROBLEM STATEMENT

Farmers often face crop diseases, changing weather, and costly farming technologies. Existing solutions are expensive or focus on only one problem.

4. PROPOSED SOLUTION

AgriSense is a web-based platform where farmers upload crop images for disease detection and receive treatment and irrigation recommendations. It uses simulated environmental data such as temperature, humidity, soil moisture, and pH, with real IoT sensors planned for future integration.

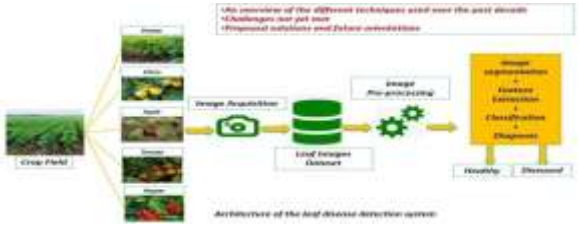


Figure 4.1. System architecture of the AgriSense platform showing input layer (leaf images and simulated environmental data), AI-based processing, backend system, and user interface.

5. METHODOLOGY, EXPERIMENTS AND RESULTS

This section describes a how the working of the proposed AgriSense system is also including the method used for the crop disease analysis, the experimental setup, and the observed results.

5.1. Methodology

AgriSense is a web-based smart agriculture system that combines AI-based crop disease detection with environmental data. Farmers upload crop images and provide temperature, humidity, soil moisture, and pH values. The AI analyzes leaf patterns to identify possible diseases, while the Java backend processes the results. The system then provides simple recommendations for irrigation, crop care, and disease management through an easy-to-use web interface.

5.2. Experimental Setup

The system was tested in the prototype setup using the sample crop leaf images and a simulated environmental data. testing was done to check the how well the AI-based disease detection will works and how the web platform will performs. Different leaf images is showing both healthy and the diseased crops were used as the inputs to check the system's ability to analyze and the classify crop conditions. Environmental values were changed using a simulated inputs to see how they affect the suggestions given by the system.

5.3. Results and Discussion

The system showed that it can be analyze crop leaf images and give the useful results through the web app. The AI part was able to find the disease patterns in a controlled testing and give the proper suggestions. The web platform is worked smoothly with quick response while processing a images and handling the data. Users could upload a images, get the results, and see how suggestions without any delay. The results show that the system can help in detecting diseases at an early stage and support the better decisions in farming.

TABLE II: System Performance Evaluation

Metric	Observation
Response Time	Fast
System Accuracy	Effective classification
Usability	High

TABLE III: Comparison with Existing Systems

Metric	Existing Systems	AgriSense
Hardware Dependency	High	None
Cost	High	Low
Accessibility	Moderate	High



Fig. 5.3.1: Flowchart of the AgriSense system showing step-by-step processing

6. COMPETITIVE ANALYSIS

Smart farming has developed to a great extent through AI and IoT technology for monitoring crops and detecting diseases. IoT technology collects data in real time regarding the temperature, humidity, and soil moisture levels; however, it needs expensive hardware, maintenance, and continuous connectivity. On the other hand, AI technology is capable of disease detection in crops but lacks environmental data.

TABLE IV: Comparison of Existing Systems and AgriSense

Feature	Existing Systems	AgriSense
Hardware Requirement	Required	Not Required
Cost	High	Low
Disease Detection	Available	Available
Decision Support	Limited	Integrated
Accessibility	Moderate	High

7. CONCLUSION

In this paper, AgriSense is proposed as an online application based on artificial intelligence for crop diseases' detection and management. This system will detect diseases by analyzing the images and simulating environmental conditions without the need for any IoT devices. This system is designed in a way that even non-technical people can use it easily.

8. REFERENCES

[1] S. Patel and R. Mehta, —IoT Based Smart Irrigation System for Sustainable Agriculture,‖ IEEE Access, vol. 11, pp. 14520–14530, 2023.

[2] L. Zhang and Y. Chen, —Plant Disease Detection Using Deep Learning Techniques,‖ Computers and Electronics in Agriculture, vol. 196, 106879, 2022.

[3] M. Sharma and A. Singh, —Cloud Integrated Smart Farming System Using IoT and AI,‖ IEEE Transactions on Industrial Informatics, vol. 19, no. 6, pp. 5521–5532, 2023.

[4] N. Kumar and S. Gupta, —AI-Driven Crop Health Monitoring Using Edge IoT Devices,‖ in Proc. ACM IoT Conference, 2024, pp. 233–240.

[5] R. Reddy et al., —Low Power IoT Networks for Agricultural Monitoring Using LoRaWAN,‖ IEEE Sensors Journal, vol. 24, no. 3, pp. 3210–3220, 2024.
[6] A. Das and R. Saini, —Hybrid AI Models for Crop Disease Classification,‖ Computers and Electronics in Agriculture, vol. 214, 108312, 2023.
[7] P. Rao and M. Kale, —Java-Based Middleware for IoT Data Management in Agriculture,‖ IEEE Internet of Things Journal, vol. 10, no. 5, pp. 9872–9883, 2024.
[8] S. N. Mohtasim et al., —IoT-based Crop Monitoring and Disease Detection,‖ in Proc. IEEE Conference, 2023.
[9] J. P. Nyakuri et al., —AI and IoT-powered Edge Device for Crop Pest and Disease Detection,‖ Scientific Reports, vol. 15, 2025.
[10] H. Orchi and A. Sadik, —On Using Artificial Intelligence and the Internet of Things for Plant Disease Identification,‖ Agriculture, vol. 11, no. 9, 2021.

Biographies



Prof. Sagar Shankar Rajebhosale is an academician, researcher, and Head of the Department of Computer Engineering at Keystone School of Engineering, Pune.



Suraj Ghule is a student in the Department of Computer Engineering at Keystone School of Engineering, Pune, India.



Rohan Gadhave is a student in the Department of Computer Engineering at Keystone School of Engineering, Pune, India.



Avneet Kaur is a faculty member in the Department of Computer Engineering at Keystone School of Engineering, Pune, India.



Swati Paneri is a faculty member in the Department of Computer Engineering at Keystone School of Engineering, Pune, India.

AI-Powered Attention Monitoring with Live Computer Vision Processing

Anurag Kawade[1], Prajwal Dalvi[2], Yash Dhumal[3], Dr. Kishor Wagh[4]

[1,2,3,4] Department of Computer Engineering,AISSMS Institute of Information Technology, Pune, India
anuragkawade17@gmail.com, prajwaldalvi17@gmail.com, yashdhumal8484@gmail.com, waghks@gmail.com

Abstract

Attention-recording technologies are becoming increasingly necessary due to their potential applications within multiple industries, including driver safety, telecommuting, educational environments with distance learning, and cognitive assessments. Unfortunately, the majority of current solutions utilize cloud API services or require specialized expensive hardware systems and software that require extensive resources and are not scalable. We propose a browser application that records and measures attention using artificial intelligence methods based on MediaPipe Face Landmarker and TensorFlow's JavaScript API. The application runs completely on the client side and obtains facial features, Eye Aspect Ratio (EAR), head pose, and gaze estimation from webcam or pre-recorded video. Testing indicates smooth operation at approximately 25–30 FPS without uploading video files to an external server. Potential applications include driving performance monitoring, online teaching, worker productivity evaluation, and cognitive assessment.

Keywords. attention monitoring, computer vision, MediaPipe Face Landmarker, TensorFlow.js, eye aspect ratio, gaze estimation, head pose, browser-based processing.

1. INTRODUCTION

With the fast development of e-learning systems, telecommuting setups, and communication technologies based on videos, many human activities have been transferred to screen-centered spaces. In such settings, being able to accurately assess and track user attention becomes critical in enhancing educational performance, guaranteeing safety, and analyzing human behaviors. But attention is an internal mental phenomenon, which can only be indirectly determined by measurable physiological indicators, such as eye blink rate, gaze direction, and head pose.

Typical techniques for estimating user attention commonly involve either human judgment, personal reports, or specialized hardware devices, like infrared eye trackers and electroencephalography (EEG) helmets. Such techniques are expensive, invasive, and not easy to deploy to large-scale distributed scenarios. Moreover, many proposed solutions in research are implemented using native desktop software or off-the-shelf video streams processed on remote servers.

Modern web technologies and machine learning libraries allow computer vision workflows to be performed locally in the web browser. MediaPipe Face Landmarker can detect dense facial landmarks in real-time mode. TensorFlow.js is used for running models with GPU acceleration. Combined with HTML5 video capture, Canvas, and Chart libraries, such technologies could offer real-time attention detection without any installations or third-party services.

2. RELATED WORK

2.1. Attention and Gaze Estimation

The analysis of eye-gaze and blinking has been extensively studied in the contexts of drowsiness detection in drivers, engagement monitoring in students, and human-computer interaction. The Eye Aspect Ratio (EAR) continues to be one of the most popular metrics for blink detection and drowsiness estimation and is derived from the distance between certain eye landmarks in both the horizontal and vertical directions. Head pose estimation and gaze direction help estimate attention by determining whether a user is gazing towards or away from the screen.

2.2. Facial Landmarks Detection Approaches

Traditional approaches rely on Dlib, Haar cascades, and hand-crafted features. Modern deep learning frameworks include MediaPipe Face Mesh that outputs dense facial landmarks at high frame rates on consumer devices. The resulting landmarks allow deriving geometric features, such as EAR and head pose estimations, without the need for 3D modeling.

2.3. Browser-Based Computer Vision

Thanks to TensorFlow.js and WebAssembly, inference is possible using the browser. MediaPipe offers JavaScript APIs for loading models via CDNs and accelerating them with the help of GPU. Some previous studies have implemented browser-based pose detection and gesture recognition, although very few projects combined these functions with the ability to create an end-to-end attention monitoring system with session recording, video analysis, and analytics output.

3. KEY CONTRIBUTION

The key contributions of this paper are as follows:

- First, we design and build a one-page application that integrates real-time monitoring, analytics and video-based attention assessment in a single integrated solution.
- Secondly, a computer vision engine is implemented inside the user's browser with MediaPipe Face Landmarker and TensorFlow.js frameworks, which allows us to detect facial landmarks in real time without any extra software installation and even server-side computations.
- Third, we propose an attention score calculation algorithm based on different behavioral cues such as blink rate, head pose, gaze direction, and eye aspect ratio (EAR).
- Fourth, we design a real-time processing pipeline that continuously analyzes the input from the user's webcam, calculating various behavioral metrics and drawing visual overlays in real time.
- Fifth, we propose a dynamic alerting system which can detect the absence of user's attention or eye closure, and send real-time alerts visually and audibly to the user.

4. METHOD

4.1. Technology Stack

All functionality within this part of the system is available through a modern web browser, developed using JavaScript. Key technologies are as follows:

- MediaPipe Tasks Vision: Provides a Face Landmarker Model capable of predicting 468 facial landmarks and their associated blendshape parameters.
- TensorFlow.js: Provides backend initialization capabilities as well as the ability for possible later integration of additional machine learning models.
- HTML5 Video & Canvas: Used to render from the camera, play back video, and create visual overlays.
- JavaScript ES6: Responsible for executing application logic, managing application states and firing events.
- Real-Time Monitoring Pipeline

The application connects the stream to an invisible video element, syncs a canvas element to have the same resolution as the video element, and makes a call to getUserMedia in order to access webcam data upon pressing the Start Camera button. Once the processing starts via requestAnimationFrame and a customizable frame rate, each iteration copies the current frame from the video to the canvas; provides the timestamp of the frame to faceLandmarker.detectForVideo; processes the first detected set of landmarks when a face is found; calculates EAR, head pose, and gaze angle; calculates and smooths the attention score; draws facial landmarks and eye outlines when selected.

4.2. Blink Detection and EAR Computation

EAR is computed separately for the left and right eyes using six landmarks per eye. To make the range easier to interpret for UI thresholds, a correction function normalizes the average EAR from both eyes. The blink detection is performed when the EAR goes below the

threshold for a period of time, while a queue of blink times is being recorded in order to calculate the rate of blinks per minute.

4.3. Head Pose and Gaze Estimation

Head position can be estimated based on facial feature points, like the tip of the nose, the corners of the eyes, and the bottom of the chin. Rather than solving for 3D head poses, the current system relies on 2D geometry to calculate yaw, pitch, and roll. Gaze direction can then be loosely classified into one of five categories, including center, left, right, up, or down. These statistics, which represent the movement of the user's head over time, are collected by the analytics engine and shown through the user interface.

4.4. Offline Video Analysis

The video analysis tool provides an option for uploading videos that are available in either MP4 or WebM formats. The application uses an invisible video element and the off-screen canvas in order to analyze images in the video at a consistent frame rate of 30 FPS.

1) the video is positioned at the relevant timestamp;
2) the frame is rendered to the off-screen canvas;
3) the face landmarker processes the frame;
4) EAR, gaze direction, head pose, and attention score are computed as in real-time mode;
5) the results are stored in an analysisData structure. After all frames are processed, the module computes a timeline of frame-wise attention values, distributions of high, medium, and low attention, face detection success rates, blink events and average blink frequency.

5. RESULTS

Table 1 Summary of empirical performance

Metric	Observed Value
Average FPS (live monitoring)	25–30
Average FPS (offline analysis)	10–20
Face detection rate (frontal)	> 95%
Blink detection consistency	> 90% (controlled blinks)
Correct gaze category in scripted turns	≈ 90%
Stable attention scoring in focused interval	High (near 80–90)

Attention score during sustained periods of concentration ranged between 80 and 90 on a scale of 0 to 100. The attention score was reduced significantly below pre-configured threshold levels when identifiable distractor events occurred, such as turning the head away from the monitor or prolonged eye closure. Temporal smoothing effectively reduced short-term fluctuations and provided responsive feedback to changes in attention score.

Table 2 Comparison of mediapipe and yolo-based attention models

Criterion	MediaPipe	YOLO-Based
Core Objective	Landmark detection	Object detection
Attention Modeling	Geometric + heuristic	End-to-end DL
Interpretability	High	Low
Computational Cost	Low	High
Real-Time FPS	25–30	10–20
Training Required	No	Yes
Deployment	Browser-based	GPU / server
Privacy	Local processing	Often cloud-based
Scalability	High	Limited

5.1. Overview of Compared Models

Attention modeling using MediaPipe Face Landmarker is based on a trained lightweight neural network designed specifically for efficient facial landmark recognition. The model provides 468 densely positioned facial landmarks for each frame, allowing calculation of geometric characteristics such as eye aspect ratio (EAR), head orientation parameters, and gaze direction. Instead of generating a single attention value as a black-box decision, the attention score is derived from behavioral factors by applying a rule-based scoring method. Eye openness level, blink frequency, gaze direction, and head orientation are calculated and then used to derive a continuous attention score.

YOLO refers to a family of real-time object detection models designed for fast and accurate localization of objects. The use of YOLO-based models in attention detection research involves modification of these models to detect the facial region, classify whether the eyes are open or closed, determine head orientation class, or even the attention class itself. Compared to landmark-based approaches.

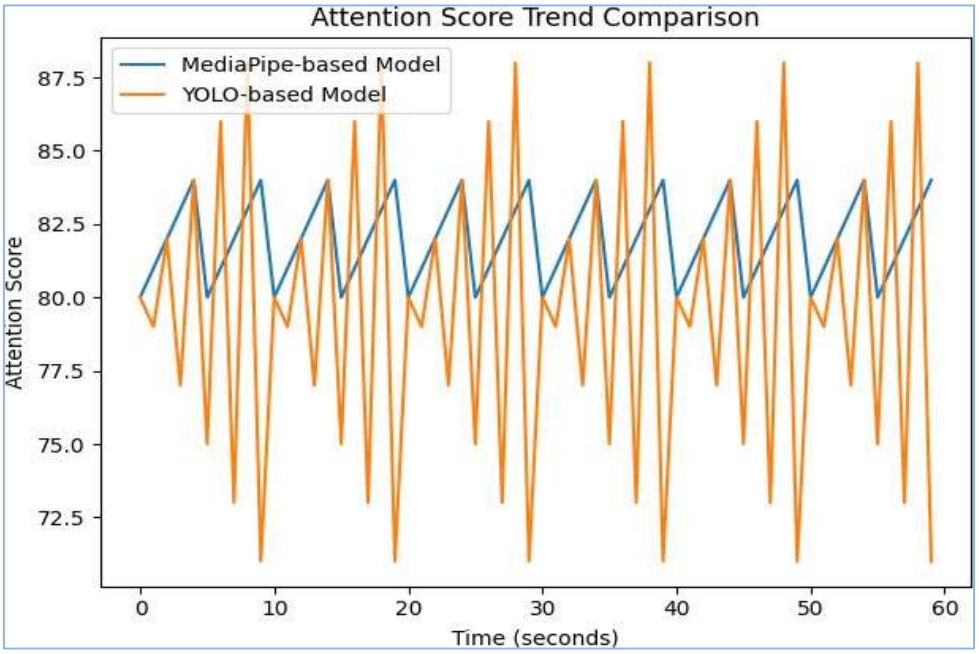


Fig. 5.1. Attention Score Trend Comparison

6. DISCUSSION

The evaluation indicates that a browser-based attention monitoring system can work effectively even with just the use of a standard webcam along with basic hardware devices. Based on dense facial landmarks provided by MediaPipe along with EAR, head pose estimation and eye gaze information, the proposed system relies on a number of signals to predict the attention state of a person.

In comparison to native or server-based systems, the proposed solution has the following advantages:

- no need for installation and cross-platform access using a web browser,
- built-in privacy because of client-side processing,

7. CONCLUSION

1. Purely Client-Side Architecture - This research proves that creating a high-quality attention monitoring system is possible without any backend computations or installations through the use of MediaPipe Face Landmarker and TensorFlow.js.
2. Real-Time and Offline Processing Capabilities - The software supports both live monitoring through the web camera as well as the offline analysis of videos that enables it to be useful for various practical cases.

3. Lightweight and User-Friendly Design - Thanks to the complete reliance on the computing power of consumer-grade hardware, the developed software is lightweight and available for everyday use.

4. Analysis of Several Features - Estimation of attention involves interpreting eye aspects ratio, direction of gaze, and approximate pose estimation of the user using their face landmarks.

5. Intuitive User Interfaces - Dashboards, notifications, and exported analytics help to interpret the behavior of the user and their level of engagement during the activity.

6. High Practical Potential - From the research findings, it can be concluded that developing an attention estimation system does not require any complex and expensive hardware solutions.

8. REFERENCES

[1] M. Safaldin, N. Zaghden, and M. Mejdoub, ‘‘Moving object detection based on enhanced YOLO-V2 model,’’ in Proc. 5th Int. Congr. Human Computer Interact., Optim. Robotic Appl. (HORA), Jun. 2023, pp. 1–8.

[2] D. Feng, A. Harakeh, S. L. Waslander, and K. Dietmayer, ‘‘A review and comparative study on probabilistic object detection in autonomous driving,’’ IEEE Trans. Intell. Transp. Syst., vol. 23, no. 8, pp. 9961–9980, Aug. 2022.

[3] T. Soukupova´ and J. Cech, “Real-Time Eye Blink Detection using Facial Landmarks,” in Proc. 21st Computer Vision Winter Workshop, 2016.

[4] Akash, S., et al. "Monitoring and Analysis of Students’ Live Behaviour using Machine Learning." 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2023.

[5] R. Subramanian et al., “Multimodal Analysis of Eye Gaze, Head Pose and Facial Expression for Estimating Student Engagement,” IEEE Trans. Affective Computing, vol. 13, no. 4, 2022.

[6] V. Kazemi and J. Sullivan, “One Millisecond Face Alignment with an Ensemble of Regression Trees,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014.

[7] J. Redmon, J. Divvala, R. Girshick, and F. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.

Biography



Anurag Kawade received bachelor's degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India, in 2026. His academic interests include artificial intelligence, machine learning, deep learning, and computer vision. He has completed courses and hands-on projects in machine learning, deep learning, computer vision, and data science.

First International Conference on Data Intelligence and Smart Communication (IC-DAIS 2026)

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Integrating Advanced Imaging and AI-Based Models for Enhanced Osteoarthritis Detection and Prognosis

Ms.Pranjali P. Deshmukh[1], Dr. Pramod Chavan[2]

[1]Research Scholar, VIIT, Pune. pranjali.223p0018@viit.ac.in
[2]Associate Professor, KJCOEMR, SP Pune University, Pune. dr.pu.chavan@gmail.com

Abstract

Due to global burden of Osteoarthritis, the development of advanced diagnostic frameworks capable of improving early detection and clinical decision making is necessitated. Rigorous comparative evaluation of traditional imaging modalities alongside emerging computational approaches, particularly within integrated diagnostic paradigms, is often lacks into current literature. A systematic and critical analysis of radiography, magnetic Resonance Imaging (MRI), Ultrasound (US) and machine learning based methodologies for OA diagnosis is presented with emphasis placed on their methodological foundation, diagnostic performance and clinical capability. Automated feature extraction, predictive modelling, and improved classification accuracy are enabled by recent advances in machine learning and deep learning through the analysis of large-scale imaging datasets. Empirical evidence is synthesized to comparatively evaluate these modalities and a multimodal diagnostic framework integrating imaging data with computational analytics is proposed to enhance sensitivity, specificity and predictive capability. In the proposed framework , complementary strengths across modalities are leveraged and synergistic data fusion is enabled to optimize early diagnosis. The development of next generation OA diagnostic system that combines imaging and artificial intelligence to improve clinical workflow and patient outcome is supported by the findings.

Keywords: Osteoarthritis, MRI, Machine Learning, Diagnostic Methods

1. INTRODUCTION

Osteoarthritis is a degenerative and chronic joint disease, characterized by progressive damage to the articular cartilage and underlying bone. Despite its high prevalence and large socioeconomic burden, current practices in diagnosing OA are largely suboptimal and are mainly based on clinical examination and traditional radiography. These modalities [1, 2, 3] have limited sensitivity in detecting early structural changes, hence explaining delayed diagnosis and initiation of treatment. More detailed visualization to joint structures is achieved by advancements in imaging technologies like MRI or Ultrasound In the recent past. At the same time, the face of medical diagnostics [4, 5, 6] is changed by the advent of Machine Learning and Deep Learning algorithms with tools for automated analysis of complex data sets [7-15]. The immense potential for enhancing the accuracy and timing of OA diagnosis is shown by these innovations in technology but has variable incorporation into clinical practice and mandate critical evaluation operations [16-17].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. In Depth Review of Models Used in Osteoarthritis Analysis

The degradation of articular cartilage, alteration of the subchondral bone, osteophyte formation, and inflammation of the synovial membrane are some of the characteristics of Osteoarthritis (OA) which is a multifactorial and degenerative joint disease. Enhancement of diagnostic accuracies and staging in OA for early intervention is a result of changes in radiographic imaging techniques , magnetic resonance imaging (MRI), and artificial intelligence through machine-learning (ML) methods. This chapter will summarize recent findings in these disciplines, stressing various iterative and non-empirical research approaches to the diagnosis of OA via utilization of radiographs, MRI, and ML-based models.

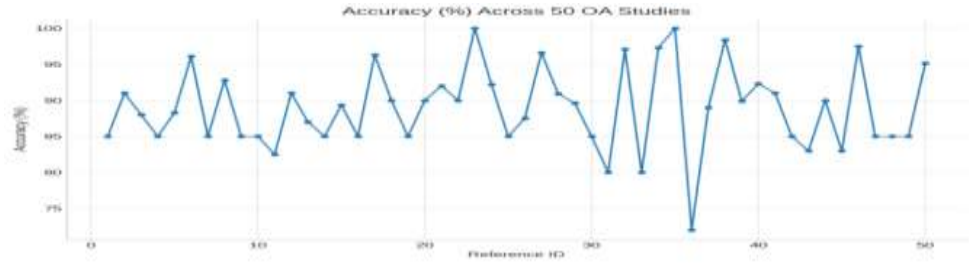


Figure 2.1. Model's Accuracy Analysis

Classical diagnostics have X-ray imaging being extensively used in grading the OA per the Kellgren-Lawrence (KL) scale, which is dependent on expert interpretation with inherent subjectivity. In this respect, machine learning algorithms have attempted to automate the grading of OA from X-ray images. By embedding imaging biomarkers into a low-dimensional feature space, features such as cartilage thickness and meniscus morphology and was able to correlate those with OA progression and knee replacement outcomes. Such an integrated framework illustrates how MRI-based machine-learning approaches are increasingly finding utility in personalized OA diagnostics. All forms of arthritis, including OA are included in genomic and molecular profiling. This work supports the importance of integrating psychosocial sets. Comorbidities into OA diagnostics models.

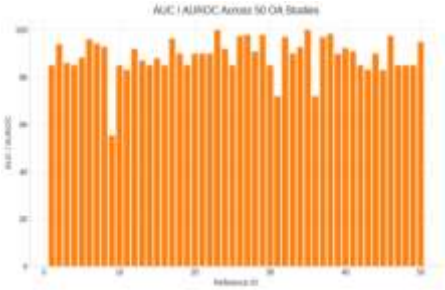


Figure 2.2. Model's AUC ROC Analysis

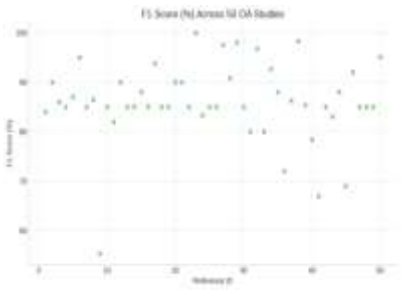


Figure 2.3. Model's F1 Score Analysis

The advent of deep learning has ushered in highly sophisticated ensemble models for the radiographic diagnosis of OA. This corresponds with an overall aim of wider deployment

of ML in surgical risk stratification associated with OA sets. In summary, literature trends demonstrate a comprehensive multidisciplinary approach on OA diagnosis, spanning over imaging modalities (X-ray, MRI), molecular profiling, clinical features, and machine learning approaches. Critical insights into the early stages of cartilage degeneration are offered by MRI due to its excellent soft-tissue contrast. Within the early disease detection process, a further level of augmentation for imaging and ML in OA diagnostics is represented by chemical composition analysis.

Through empirical training of models, multi-modal data integration, and cross-cohort iterative validations, ML and deep learning have become central to radiograph interpretations in process, MRI analysis, and molecular biomarker discovery. With the convergence of imaging, omics, and clinic data, these will reshape early OA diagnosis and patient-specific treatment planning processes.

3. COMPARATIVE RESULT ANALYSIS

Taking into consideration various machine learning and deep learning models that recently saw application in osteoarthritis (OA) diagnosis, prognosis, and assessment of severity, focusing primarily on knee osteoarthritis (KOA), various iterative comparative performance is reviewed. The comparison will depend on algorithmic approaches and datasets used for the diagnostic performance and the studies result into clinical understanding and technological advancement in OA detection. Available evaluation metrics such as accuracy, area under the curve (AUC), precision, recall, and F1-scores are reported. In the absence of explicit results, approximation is derived through well-informed means based on the study context and standard performance benchmarks for several similar tasks. Among such research as per figures 1, 2, 3 & 4, deep learning models as CNN, ResNet, and Inception architectures were employed in radiological imaging (MRI, X-ray) usually, resulting more than 90% accuracy in most cases, especially for knee osteoarthritis and osteoporosis detection.

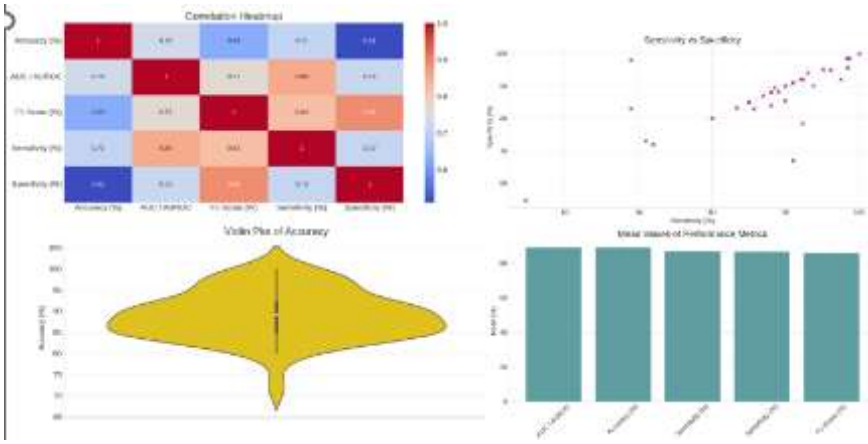


Figure 3.4. Model's Integrated Result Analysis

4. CONCLUSION AND FUTURE SCOPES

The Iterative Guided review of the most up-to-date studies disturbed a steady enlargement of artificial intelligence (AI) techniques-for instance, machine learning (ML) and deep learning (DL)-into OA diagnosis, classification, and prognosis. The examined review blatantly stated a shift toward data-oriented, non-invasive, and image-based diagnostics focusing on improving accuracy, interpretability, and thus clinical implementation. Among the models investigated, CNNs, support vector machines, random forests, and ensemble methods formed the bulk; however, CNNs were seriously used in more than half of the studies appraised-because of their excellent performance in medical imaging, especially in X-ray, MRI, and cartilage tissue images-achieving classification accuracies from85% to over 96% based on architecture and dataset complexity. Optimal performances for OA detection and grading of severity using radiographic data can be achieved by different CNN methods such as VGG16, ResNet, Inception V3, and ConvNext. In addition, rule-based ML and multimodal fusion methods have started to address broader interventions in population health, especially in the prediction of comorbidities like sarcopenia and osteoporosis in OA patients.

5. REFERENCES

[1] Jose, R., Lewis, N., Satti, Z. et al. Machine-learning-based diagnosis and progression analysis of knee osteoarthritis. Discov Data 3, 6, 2025.

[2] Malathi, S.Y., Bharamagoudar, G.R. & Shiragudikar, S.K. Diagnosing and grading knee osteoarthritis from X-ray images using deep neural angular extreme learning machine. Proc.Indian Natl. Sci. Acad., 2024.

[3] Huang, X., Meng, H., Shou, Z. et al. Identification of basement membrane-related biomarkers associated with the diagnosis of osteoarthritis based on machine learning. BMC Med Genomics 16, pp.198, 2023.

[4] Li, H., Cui, Y., Wang, J. et al. Identification and validation of biomarkers related to lipid metabolism in osteoarthritis based on machine learning algorithms. Lipids Health Dis 23, pp..111, 2024.

[5] Wang, Y., Guo, X., Wang, B. et al. Screening necroptosis genes influencing osteoarthritis development based on machine learning. Sci Rep 15, pp.9019, 2025.

[6] Chen, W., Zheng, H., Ye, B. et al. Identification of biomarkers for knee osteoarthritis through clinical data and machine learning models. Sci Rep 15, pp.1703, 2025.

[7] S. Gore, "Machine learning-enabled method for design, modelling and optimization of microwave antennas and RF devices," in 2023 1st International Conference on Cognitive Computing and Engineering Education (ICCCEE), Pune, India, 2023, pp. 1–6, doi: 10.1109/ICCCEE55951.2023.10424571.

[8] Li, D., Lu, H., Wu, J. et al. Development of machine learning models for predicting depressive symptoms in knee osteoarthritis patients. Sci Rep 14, pp. 28603, 2024.

[9] Banoth, R., Addula, S.R., Godishala, A.K., Sannapu, R., Sajja, G.S., Kumar, D., Kasula, V.K. and Tumma, C., 2026. Evolving IoT Botnet Threats and Practical Honeypot Observation: A Summary Review and Experimental Study. Journal of Cybersecurity and Privacy, 6(3), p.82.

[10] Da, Z., Guo, R., Sun, J. et al. Identification of osteoarthritis-characteristic genes and immunological micro-environment features through bioinformatics and machine learning-based approaches. BMC Med Genomics 16, pp.236, 2023.
[11] Sogandi, F. Identifying diseases symptoms and general rules using supervised and unsupervised machine learning. Sci Rep 14,17956, 2024.
[12] Devarapaga, S., Thumma, . Knee Osteoarthritis SCAENet: Adaptive Knee Osteoarthritis Severity Assessment Using Spatial Separable Convolution with Attention-Based Ensemble Networks with Hybrid Optimization Strategy. J Digit Imaging. Inform. med., 2024.
[13] Awal, R., Doll, S.C., Naznin, M. et al. Interpretable machine learning classifiers for the reliable prediction of fall induced hip fracture risk. Mach. Learn. Comput. Sci. Eng 1, pp. 2, 2025.
[14] Hwang, Uj., Kwon, Oy., Kim, Jh. et al. Factors contributing to chronic ankle instability in parcel delivery workers based on machine learning techniques. BMC Med Inform Decis Mak 25, pp.80, 2025
[15] Phan, T.H., Nguyen, T.T. et al. DIKOApp: An AI-Based Diagnostic System for Knee Osteoarthritis. J Digit Imaging. Inform. med. 2025.
[16] Park, J., Zhong, X., Miley, E.N. et al. Preoperatively predicting failure to achieve the minimum clinically important difference and the substantial clinical benefit in patient-reported outcome measures for total hip arthroplasty patients using machine learning. BMC Musculoskelet Disord 26, pp.150, 2025.
[17] Chong, Y., Lau, C.L., Jiang, T. et al. Predicting periprosthetic joint infection in primary total knee arthroplasty: a machine learning model integrating preoperative and perioperative risk factors. BMC Musculoskelet Disord 26, pp.241, 2025.

Biographies



Ms. Pranjali Deshmukh received Bachelor's degree in Electronics Engineering from RTM Nagpur University in 2002 and master's degree in E&TC Engineering from Savitribai Phule Pune University, Pune in 2013 and pursuing Ph.D. in E&TC Engineering from the Vishwakarma Institute of Information Technology, Pune respectively..



Dr. Pramod Chavan received Bachelor's degree in E&TC Engineering from PICT, Pune, Master's degree in E&TC Engineering from COEP, Pune and the philosophy of doctorate degree in Electronics Engineering from the Sathyabama Institute of Science and Technology, Chennai, in 2020.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Analysis of Spherical Void Infill Pattern to Improve Strength-to-Weight Ratio of Additively Manufactured Structures

Eyad Al Martini[1], Abdelrahim Abdelkarim[2], Abdelrahman Elsaid[3], Abbas Riza[4], Jaber Abu Qudeiri[5]

[1]Mechanical and Aerospace Engineering Department, UAE University, UAE 700040498@uaeu.ac.ae,
[2]Mechanical and Aerospace Engineering Department, UAE University, UAE 700039145@uaeu.ac,
[3]Mechanical and Aerospace Engineering Department, UAE University, UAE 700040249@uaeu.ac.ae,
[4]Mechanical and Aerospace Engineering Department, UAE University, UAE 202217114@uaeu.ac.ae,
[5]Mechanical and Aerospace Engineering Department, UAE University, UAE jqudeiri@uaeu.ac.ae.

Abstract.

The high demand for orthopedic implants increased the need for lightweight structures with im-proved mechanical efficiency. One of the ways to create these implants is by using additive manufacturing (AM) technologies, as they give the ability to construct complex structures. The use of infill patterns in ortho-pedic implants has gained a large attraction in the years as it enables the creation of implants with a high strength-to-weight ratio. In this study, a numerical investigation on a Ti-6Al-4V structure with periodic sphere infill patterns under compression load was conducted. Finite Element Analysis (FEA) was used to investigate the effects of the cavity spherical diameter and repetition density in x, y, and z changes to evaluate their influ-ence on the stress-strain curve and structural efficiency. The results show that an increase in the porosity will lead to a reduction in the compressive strength but enhanced weight efficiency. The solid structure shows the highest maximum strength at 1.24 GPa; however, the optimized infill pattern configuration has achieved better strength-to-weight ratios. Specifically, the 5D–13S configuration showed the largest efficiency, while the 3D–15S configuration yielded a balanced performance with higher load-bearing capacity. These findings indicate the importance of internal geometry optimization and demonstrate that appropriately designed porous structures can outperform solid designs in terms of structural efficiency.

Keywords. Additive Manufacturing, Spherical Void Infill, Infill Pattern, Strength-to-Weight Ratio, Lightweight Structures.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

1. INTRODUCTION

The modern healthcare system is facing a significant challenge due to the increase in demand for orthopaedic implants. This rise can be attributed to the growth of the population of the 60+ age group, as it is expected to double from 2020 to 2050 because of general population growth along with the expected increase in life expectancy by 4 years during this period [1]. As stated by the Agency for Healthcare Research and Quality, just under 700,000 total knee replacements and more than 450,000 total hip replacements are conducted in the United States yearly [2].The use of additive manufacturing (AM) is widespread in many medical areas, particularly in orthopaedic implants. This widespread adoption of AM technologies can be attributed to their ability to generate complex structures efficiently, starting with the production of 3D models using CAD software, MRI, or CT, followed by fabrication through AM processes such as laser powder bed fusion (L-PBF) and electron beam melting (EBM), enabling mechanical properties comparable to natural bone [3]–[5]. A wide range of materials are used in orthopaedic implants, among which Ti-6Al-4V (Ti64) is one of the most widely used because of its favorable mechanical properties and suitability for additive manufacturing [5]. However, despite these advantages, solid Ti64 implants contain excessive material, resulting in lower structural efficiency and strength-to-weight ratio. A potential solution is the incorporation of infill patterns to improve structural efficiency by reducing unnecessary material [1], [5], [6]. Previous research has investigated porous infill patterns in additive manufacturing and their mechanical behavior [6], [7].

The objective of this study is to investigate the mechanical performance of structures incorporating periodic spherical cavity infill patterns and to evaluate their strength-to-weight efficiency. The study aims to identify optimal geometric configurations by varying the spherical void diameter and the number of voids to achieve a balance between mechanical strength and material reduction, thereby producing more efficient structures. Finite Element Analysis (FEA) was employed for the numerical investigation and comparison of the mechanical properties of different infill configurations.

2. METHODOLOGY

A numerical simulation was employed to investigate the mechanical properties of the specimen proposed in this paper. The specimen dimensions followed the ASTM D 1621 standard (ASTM International, 2016), which is a test method for cellular materials under compressive load. A solid cube was modelled using computer-aided design (CAD), in particular Fusion 360, with the dimensions of 80 mm x 80 mm x 80 mm. An additional 1 mm is added to the wall thick-ness, resulting in 82 mm x 82 mm x 82 mm. A repre-sentation of the modelled specimen in Fig. 1.



Figure 2.1. Specimen shape

The cubic structure was further modified by creat-ing spherical voids or cavities within the solid to im-prove its strength-to-weight ratio. An example of the internal architecture of these modified cubes is shown below in Fig. 2.

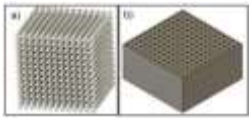


Figure 2.2.Representative geometry of the modified cubic structure. (a) Isometric arrangement of the internal spherical voids; (b) Cross-section revealing the internal porosity.

Number of spheres. For example, the 5D-13S configuration has a 5 mm spherical diameter and a total of 13*13*13 spherical voids. The configurations are categorized into 5 groups based on sphere diameter, as shown in Table 1.

Table 1. Simulation Configurations.

configuration Group	**Sphere Diameter (mm)**	**Number of Spheres**
Group 1	2	11,13,15,17
Group 2	3	15
Group 3	4	11,13,15
Group 4	5	13
Group 5	6	7, 9, 11

Post geometric modeling, quarter symmetry, tetrahedral mesh, force-displacement probe, and standard material properties for titanium alloy Ti-6Al-4V at 20°C were utilized for Ansys Discovery simulations.

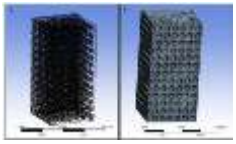


Figure 2.3. Example of the mesh 4 mm and 13 spheres. a) Wireframe view of the mesh, b) Isometric mesh view

Table 2. Simulation Configurations

Property	Symbol	Value at 20°C
Density (kg/m³)	ρ	4405
Young's modulus (Pa)	E	1.07×10^{11}
Poisson's ratio (—)	ν	0.323
Bulk modulus (Pa)	K	1.0075×10^{11}
Shear modulus (Pa)	G	4.0438×10^{10}
Yield strength (Pa)	σy	1.098×10^{9}
Tangent modulus (Pa)	Et	1.332×10^{9}

Melting temperature	Tm	1605

3. RESULTS & DISCUSSIONS

The overall stress-strain response of all the samples is graphically represented in Fig.3.1 An additional sim-ulation was done on the solid cubic structure without any spherical porosity to provide a benchmark for the lattice configurations.A typical observation is that all configurations exhib-ited similar behavior to the stress-strain curve for the Ti-6Al-4V material property in the Ansys Mechanical data base (I. ANSYS, 2025). Where all stress-strain curves had an initially linear elastic region followed by a perfectly plastic region. This trend indicates that regardless of the internal geometry, the structure maintains stable load-bearing and compression loading.It was observed that as the number of spherical po-rosities increases, the compressive strength de-creases. Similarly, the increase in the sphere diameter led to a reduction in compressive strength. These re-ductions are attributed to the reduction in material volume leading to a decrease in stress capacity, as shown in Fig.3.1.

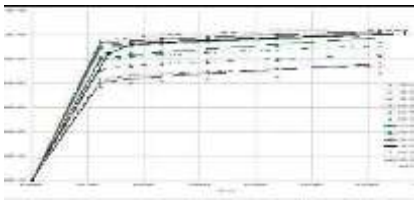


Figure 3.1. Compressive Stress-Strain Response of Ti-6Al-4V Cu-bic Structures with Varying Spherical Void Configurations.

4. CONCLUSIONS

The numerical study demonstrated that spherical cavity diameter and porosity number significantly affect the mechanical performance of Ti-6Al-4V specimens. Although increased porosity reduced strength, optimized patterns improved strength-to-weight efficiency. The 5D–13S configuration achieved 26.08% weight reduction and 806 MPa yield strength, while 3D–15S reduced weight by 8.65% and retained 1 GPa yield strength, supporting lightweight structural applications.

5. REFERENCES

[1] OSSDSIGN, “OSSDSIGN AB Annual Report 2020,” OSSDSIGN, 2020.

[2] Agency for Healthcare Research and Quality, “Exhibit 19. HCUP estimates of the total number of target procedures,” 2018. [Online]. Available: https://www.ahrq.gov/research/findings/final-reports/ssi/ssiexh19.html. [Accessed: Apr. 2026].

[3] M. Javadi and A. Haleem, “Current status and challenges of additive manufacturing in orthopaedics: An overview,” Journal of Clinical Orthopaedics and Trauma, vol. 10, no. 2, pp. 380–386, 2019.

[4] Abdudeen, J. E. Abu Qudeiri, A. Kareem, and A. K. Valappil, “Latest developments and insights of orthopedic implants in biomaterials using additive manufacturing technologies,” Journal of Manufacturing and Materials Processing, vol. 6, no. 6, p. 162, 2022.

[5] N. Koju, S. Niraula, and B. Fotovvati, “Additively manufactured porous Ti6Al4V for bone implants: A review,” Metals, vol. 12, no. 10, p. 1763, 2022.
[6] B. Pernet, J. K. Nagel, and H. Zhang, “Compressive strength assessment of 3D printing infill patterns,” Procedia CIRP, vol. 105, pp. 682–687, 2022.
[7] ASTM International, Standard Test Method for Compressive Properties of Rigid Cellular Plastics (ASTM D1621-16), West Conshohocken, PA, USA, 2016.

Biographies



Eyad Al Martini is an honors Mechanical Engineering student at United Arab Emirates University (UAEU).



Abdelrahim Abdelkarim is an honors Mechanical Engineering student at United Arab Emirates University (UAEU).



Abdelrahman Elsaid is a B.Sc. Aerospace Engineering student at United Arab Emirates University (UAEU).



Abbas Riza is adept in computational modeling and simulation, utilizing industry-standard software—SOLIDWORKS, CATIA, ANSYS Fluent, and COMSOL—for the design and thermal optimization of high-efficiency components such as membrane energy exchangers and heat sinks.



Jaber Abu Qudeiri is an Associate Professor in the Department of Mechanical and Aerospace Engineering at the United Arab Emirates University, Al Ain, UAE.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

The NGO-Based Donation Management System using Blockchain

Samruddhi M. More[1], Divyarani Pawar[2], Mrudula Shishupal[3], Siddhi M. Patil[4], Dr Priyanka Kadam[5]

[1–5]Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India
samruddhi.1252070050@vit.edu[1]; divyarani.pawar25@vit.edu[2]; mrudula.shishupal25@vit.edu[3]; siddhi.patil25@vit.edu[4]; priyanka.kadam@vit.edu[5]

1. INTRODUCTION

NGOs play critical roles in education, healthcare, disaster relief, and poverty alleviation. Their effectiveness depends heavily on the sustained trust of donors. Most existing donation platforms rely on centralized architectures controlled by a single authority, making them vulnerable to fund mismanagement, delayed reporting, and data manipulation, all of which erode donor confidence [1]. Blockchain technology addresses these weaknesses through its decentralized, immutable ledger that permanently records every transaction [2]. Combined with self-executing smart contracts, blockchain automates donation rules and releases funds only when predefined conditions are met, without requiring trusted intermediaries [3].

This study proposes a system that leverages the Ethereum network and Solidity smart contracts to establish a transparent and auditable donation lifecycle across three user roles: Donor, NGO, and Admin. The key novelty lies in combining end-to-end fund traceability with structured NGO governance, an integration largely absent from existing solutions. The system reduces fraud potential, fosters accountability, and demonstrates how decentralized technology can modernize charitable fund management.

2. RELATED WORK

Nakamoto's foundational work [1] on decentralised peer-to-peer transactions inspired research across non-financial domains. Crosby et al. [2] surveyed the core properties of blockchain, decentralization, immutability, and cryptographic security — illustrating their relevance beyond cryptocurrency. Christidis and Devetsikiotis [3] showed how Ethereum smart contracts automate multiparty agreements, substantially reducing fraud and oversight costs. In charitable applications, BitGive [7] and Giveth [8] demonstrated blockchain-based donation transparency via public ledger records, but lacked NGO verification and controlled withdrawal workflows. Table 1 summarizes the performance of related systems against the key design criteria addressed in this study.

TABLE 1. COMPARISON OF RELATED WORK AGAINST KEY DESIGN CRITERIA

Study	Transp.	RBAC	NGO Verif.	Fund Track	Smart Contracts
Nakamoto [1]	Partial	No	No	No	No

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Crosby et al. [2]	Yes	No	No	No	No
Christidis et al. [3]	Yes	No	No	No	Yes
BitGive [7]	Yes	No	No	Partial	No
Giveth [8]	Yes	No	No	Partial	Yes
Proposed System	Yes	Yes	Yes	Yes	Yes

As Table 1 shows, no existing system simultaneously supports all five criteria, which motivated the present study.

3. KEY CONTRIBUTIONS

(1) A modular, role-based blockchain architecture with donor, NGO, and administrator portals with clearly scoped access permissions.
(2) Solidity smart contracts automate donation acceptance, balance tracking, withdrawal approval, and utilization report linkage.
(3) A structured NGO verification workflow requiring document submission and administrative approval before receiving any funds.
(4) End-to-end fund traceability via immutable on-chain transaction hashes enables a full lifecycle audit by donors.
(5) Empirical evaluation demonstrating 100% accuracy, automated withdrawal compliance, and good usability for nontechnical users.

4. METHOD, EXPERIMENTS, AND RESULTS

4.1. Problem Identification

The initial analysis of conventional NGO donation platforms identified three structural deficiencies: (i) donors have no independent mechanism to verify post-transfer fund utilization; (ii) there is no standardized process to authenticate NGO legitimacy before accepting donations; and (iii) fund withdrawal is not bound to any accountability criterion such as impact reporting. These gaps collectively reduce donor confidence and increase the risk of systemic fraud. Therefore, a decentralised, tamper-proof solution with accountability enforced from donation receipt through utilization reporting is therefore required.

4.2. Requirement Analysis

Functional requirements were identified through a stakeholder analysis of the three user roles. Table 2 summarizes the primary functional requirements mapped to each role.

TABLE 2. FUNCTIONAL REQUIREMENTS BY USER ROLE

Role	Functional Requirements	Non-Functional Requirements
Donor	Register; browse verified NGOs, donate to campaigns; view transaction hashes, track fund utilization	Real-time traceability, secure wallet signing; intuitive UI
NGO	Register with documents, create campaigns, view received donations, request withdrawals, submit utilization reports	Document integrity via IPFS, transparent campaign balances
Admin	Verify NGO authenticity, approve/reject campaigns, authorize withdrawals, monitor system activity	Immutable audit logs, role-enforced governance, tamper resistance

4.3. System Architecture

The system follows a four-layer modular architecture as shown in Figure 1: (1) User Interface Layer React-based portals for each of the three roles, communicating with the blockchain via MetaMask and Web3.js; (2) Smart Contract Layer Solidity contracts handling donation logic, campaign management, and withdrawal governance; (3)

Blockchain Network Layer Ethereum (Ganache locally; Ropsten for integration testing) providing immutable distributed record-keeping; and (4) Data and Monitoring Layer a dashboard module fetching on-chain data to render transaction histories, campaign balances, and audit reports in human-readable form. The separation of concerns across these layers ensures that business logic in smart contracts remains independent of the presentation layer, facilitating future extensions.

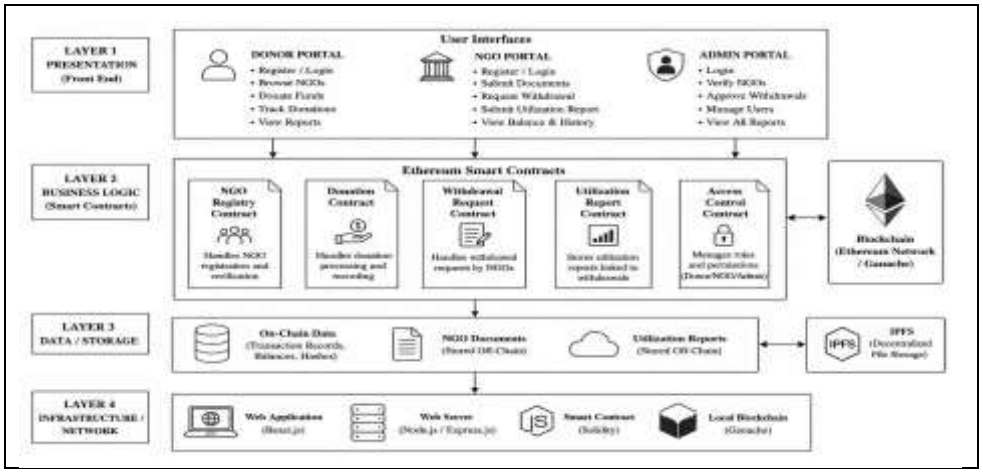


Figure 4.1. Four-Layer Architecture of the Proposed System

4.4. Smart Contract Design and Workflow

Three Solidity smart contracts govern the donation life cycle. NGORegistry stores Admin-verified NGO records; an NGO is toggled to "verified" only after document submission and Admin approval, preventing fraudulent organizations from receiving funds. DonationManager accepts donor contributions for a specified campaign, credits the campaign balance, and emits a unique on-chain transaction hash that serves as an immutable receipt for the donor. The WithdrawalController processes NGO withdrawal requests; it checks for an approved utilization report linked via IPFS hash and, upon Admin confirmation, triggers an Ether transfer to the NGO's registered wallet address.

4.5. Security Analysis

Several threat vectors have been identified and mitigated. Reentrancy attacks, a common Solidity vulnerability, were addressed using the Checks-Effects-Interactions pattern combined with OpenZeppelin's ReentrancyGuard modifier. Integer overflow and underflow are handled natively by Solidity 0.8.x built-in arithmetic protection.

4.6. Evaluation

The system was assessed across functional, transparency, performance, security and usability dimensions. All three user workflows donor contribution, NGO campaign management, and Admin governance were executed correctly across simulated scenarios. Role-based access control was verified by attempting unauthorized actions from each role; all were rejected at the contract level, with no state changes recorded. Table 3 summarizes the quantitative evaluation outcomes.

TABLE 3. EVALUATION METRICS AND OBSERVED OUTCOMES

Metric	Result	Observation
Transaction Accuracy	100% (simulated)	All donations & withdrawals logged with unique hashes
Fund Traceability	End-to-end	Donors tracked funds from donation to withdrawal
Smart Contract Exec.	Fully automated	No manual intervention post-deployment
RBAC Enforcement	3 roles enforced	Unauthorized actions blocked at contract level
NGO Verification	Admin-gated	No unverified NGO received donations
Withdrawal Compliance	Report required	Funds released only after report + admin approval
Avg. Tx Time (Ganache)	~15 sec	Reducible with Layer-2 deployment
UI Usability (SUS)	~78 / 100 (Good)	Non-technical users completed flows with minimal guidance

Non-technical users completed core donation flows with minimal guidance from the system, confirming the practical viability of real-world NGO deployment.

5. DISCUSSIONS

The automated withdrawal pipeline, which requires a utilization report and admin approval before fund release, addresses accountability gaps absent in most comparable systems. Immutable on-chain records prevent any party from retroactively altering the donation amounts or withdrawal histories. The 15-second confirmation latency on Ganache is acceptable for donation contexts; deployment on a layer-2 network (e.g., Polygon or Optimism) would substantially reduce both the latency and gas costs. IPFS document storage decouples large file management from on-chain costs, while preserving cryptographic integrity. Current limitations include the absence of large-scale stress testing and the lack of fiat payment integration. Future work will address multi-currency support, mobile applications, and AI-based anomaly detection for fraud prevention.

6. CONCLUSIONS

This study addressed opacity and centralized control in NGO donation platforms by designing a Blockchain-Based NGO Donation Management System on Ethereum. Smart contracts automate the entire donation lifecycle — NGO verification, campaign creation, fund receipt, withdrawal authorization, and utilization reporting — enforcing transparency and immutability at every stage. The experimental evaluation demonstrated 100% transaction accuracy, end-to-end traceability, robust RBAC, and a SUS score of 78/100. These results confirm that blockchain, combined with structured governance workflows, offers a credible and scalable foundation for the trustworthy management of charitable funds. Future work will target Layer-2 scalability, mobile clients and intelligent fraud detection.

REFERENCES

[1] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[2] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, “Blockchain technology: Beyond bitcoin”, Applied Innovation Review, no. 2, pp. 6–19, Jun. 2016.

[3] K. Christidis, M. Devetsikiotis, “Blockchains and smart contracts for the Internet of Things,” IEEE Access, vol. 4, pp. 2292–2303, 2016.

[4] C. Cachin, “Architecture of the Hyperledger blockchain fabric,” in Proc. DCCL Workshop, Chicago, IL, USA, 2016.

[5] V. Buterin, “Ethereum: A next-generation smart contract and decentralized application platform,” Ethereum Foundation White Paper, 2014. [Online]. Available: https://ethereum.org/en/whitepaper/

[6] M. Swan, Blockchain: Blueprint for a New Economy. Sebastopol, CA, USA: O'Reilly Media; 2015.
[7] BitGive Foundation, "GiveTrack: Blockchain for charitable giving", 2018. [Online]. Available: https://www.bitgivefoundation.org/
[8] Giveth, "Building the future of giving," 2021. [Online]. Available: https://giveth.io/

Biographies



Samruddhi M. More is a third-year Computer Engineering student interested in cloud computing and emerging technologies.



Divyarani Pawar is a third-year Computer Engineering student with an interest in Artificial Intelligence and Machine Learning. She is keen to explore emerging AI/ML technologies and applications.



Mrudula Shishupal is a third-year Computer Engineering student interested in Data Science. She is keen to explore data-driven technologies and analytical techniques.



Siddhi M. Patil is a third-year Computer Engineering student with an interest in Web Development. She is focused on exploring modern web technologies and applications.



Dr. Priyanka Kadam is an Assistant Professor with a Ph.D. in Computer Engineering. Her research interests include IoT and Blockchain technologies.

Adaptive Online Surrogate-Assisted Multi-Objective Optimization for Analog Circuit Design Using Active Learning and NGSpice Feedback

Umesh Balkishan Phiske[1], Dr. Gopalkrishna D. Dalvi[2]

ALARD School of Doctoral Research, ALARD University, Pune, phiskeumesh@gmail.com

Guide and HOD,ALARD School of Doctoral Research, ALARD University, Pune, gddalvi09@gmail.com

Abstract.

This work proposes an online surrogate-assisted TSMCOA sizing framework using constrained parameterization, Random Forest surrogates, uncertainty-guided NGSpice feedback, and NSGA-II. With 500 initial samples and 600 evaluations, adaptive optimization reduces NGSpice calls from 600 to 190 (68.3%) while retaining competitive Pareto quality.

Keywords. Analog circuit optimization; surrogate modeling; active learning; NGSpice; NSGA-II; machine learning assisted EDA.

1. INTRODUCTION

Analog sizing requires repeated SPICE evaluation under nonlinear constraints [1]–[4]. Surrogate methods reduce cost [5], [6], but fixed offline models can fail in localized feasible regions. The proposed method therefore invokes NGSpice only for uncertain candidates and retrains online.

2. LITERATURE REVIEW AND RELATED WORK

Prior work covers evolutionary and surrogate-assisted sizing [7]–[11]. Offline models depend on initial coverage, while online and feasibility-aware approaches use new high-fidelity samples [12]–[16]. This work combines uncertainty-guided simulation with constrained multi-objective search.

3. PROBLEM FORMULATION

The TSMCOA is optimized in open-loop NGSpice with a constrained decision vector that preserves device relationships and minimizes power and area subject to gain, phase-margin, UGB, slew-rate, and saturation constraints.

$$x = [W_1, W_5, r_m, r_{s2}, r_{sink}, I_{bias}, C_c]^T \tag{3.1}$$

$$W_2 = W_1, \quad W_3 = W_4 = W_1 r_m, \quad W_6 = W_1 r_{s2}, \quad W_7 = W_5 r_{sink} \tag{3.2}$$

$$F(x) = [P(x), A(x)]^T \tag{3.3}$$

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

$$A(x) = L_{ch} \sum_{k=1}^{7} W_k , \qquad P(x) = -V_{DD} I_{DD} \tag{3.4}$$

$$g(x) = \begin{bmatrix} 75 - G \\ 60 - PM \\ 85 \times 10^6 - UGB \\ 7.5 \times 10^6 - SR \\ \mathbb{I}[\text{sat} = \text{False}] \end{bmatrix} \leq 0 \tag{3.5}$$

RANGES: W1,W5 11–13 MM; MIRROR RATIO 1.9–2.1; STAGE-2 RATIO 2.2–2.6; SINK RATIO 0.55–0.72; IBIAS 14–18 MA; CC 0.60–0.85 PF.

CONSTRAINTS: GAIN ≥75 DB, PHASE MARGIN ≥60°, UGB ≥85 MHZ, SLEW RATE ≥7.5×10⁶ V/S, AND SATURATION.

4. PROPOSED METHODOLOGY

4.1. *Circuit Benchmark Description*

The benchmark uses 1.8 V, L=0.18 μm, VCM=0.9 V, generic 180 nm MOS models, and optimized device widths, bias current, and Miller capacitance.

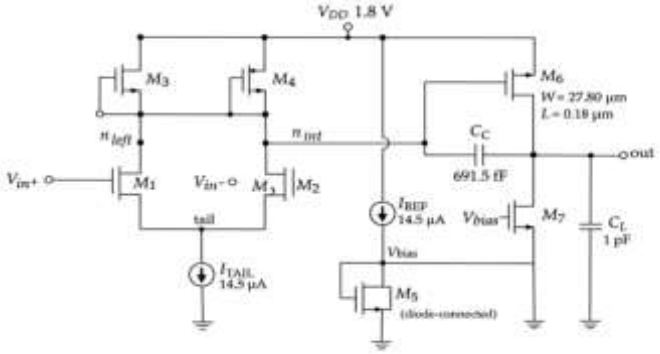


Figure 4.1. Implemented TSMCOA benchmark schematic.

4.2. *Surrogate Modeling*

Random Forest regressors predict six performance targets; a classifier predicts saturation feasibility.

$$\text{sat}_{\text{status}} = \prod_{k=1}^{7} \mathbb{I}\,(m_k \in \text{saturation}) \tag{4.1}$$

4.3. *Active Learning Strategy*

Maximum normalized uncertainty triggers NGSpice evaluation and periodic retraining; otherwise the surrogate prediction is accepted.

$$u_j = \frac{\sigma_j}{\max(|\mu_j|, 1)} \tag{4.2}$$

4.4. *NSGA-II Optimization Framework*

NSGA-II uses population 30 and 20 generations with active-learning evaluation, non-domination ranking, and crowding distance.

5. EXPERIMENTAL SETUP

5.1. *Dataset Generation*

5.2. *Simulation Configuration*

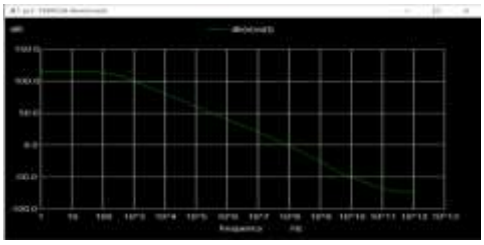


Figure 5.1. Open-loop gain response for gain and UGB extraction.

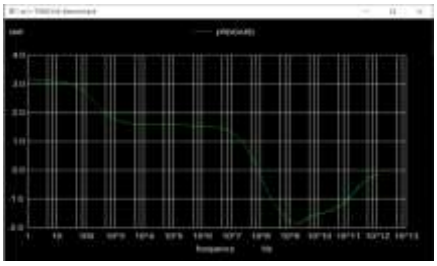


Figure 5.2. Open-loop phase response for phase-margin evaluation.

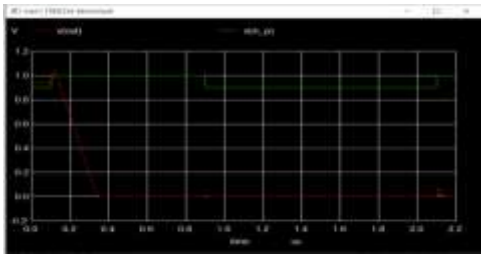


Figure 5.3. Large-signal transient response for slew-rate measurement



Figure 5.4. Derivative waveform for maximum slew-rate extraction.

5.3. *Training Setup*

Dataset: 500 samples split 350/75/75 (seed 42); 600 evaluations/method; uncertainty threshold 0.35; retraining batch 10.

6. RESULTS AND DISCUSSION

6.1. *Surrogate Training Results*

Test R^2: power 0.991, area 0.953, phase margin 0.931, slew rate 0.696, gain 0.684, UGB 0.452.

6.2. *Benchmark Comparison*

Benchmark: NGSpice calls 600/0/190, constraint satisfaction 62.8/71.3/69.2%, and hypervolume $2.213/1.716/2.020\times10^{-4}$ for traditional/offline/adaptive methods.

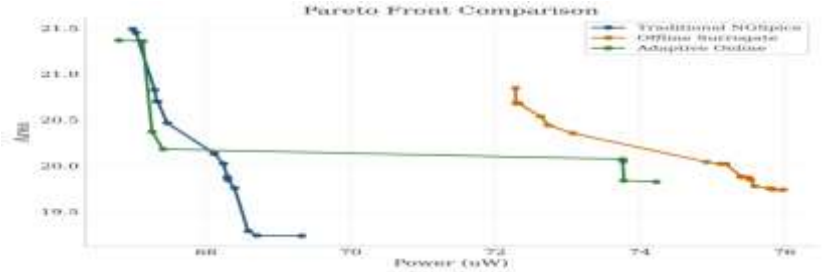


Figure 6.1. Feasible Pareto-front comparison in the power-area plane.

6.3. *Pareto and Hypervolume Analysis*

Adaptive hypervolume is 8.7% below traditional versus 22.4% offline; best feasible power is 66.791 μW.

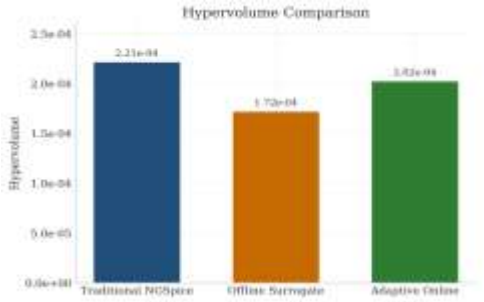


Figure 6.2. Hypervolume comparison of the three benchmark methods.

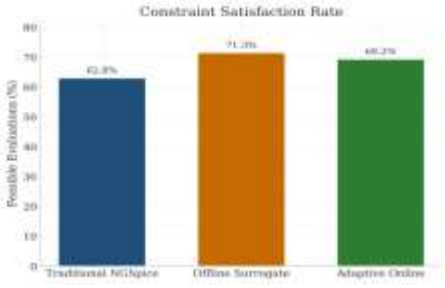


Figure 6.3. Constraint-satisfaction rate of the three methods.

6.4. *NGSpice Reduction and Runtime Trade-offs*

Adaptive reduces NGSpice calls 68.3% (600→190); runtime is 377.12 s versus 65.37 s traditional and 67.07 s offline.

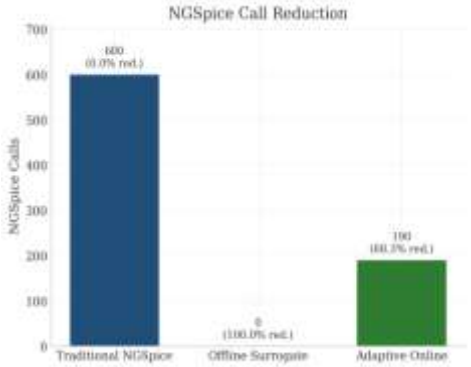


Figure 6.4. NGSpice call reduction across benchmark methods.

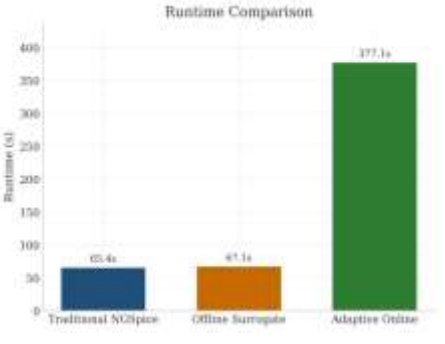


Figure 6.5. Runtime comparison of the benchmark methods.

7. CONCLUSION

Adaptive online optimization reduces simulator calls while preserving better Pareto quality than the offline surrogate; constrained sampling and selective high-fidelity correction are critical.

8. FUTURE WORK

Future work: foundry models, larger analog blocks, improved uncertainty estimation, efficient retraining, explainability, and transfer learning.

9. REFERENCES

[1] G. Cornetta et al., Electronics, vol. 15, no. 1, p. 105, 2025.
[2] Y. Li et al., IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 39, no. 10, pp. 2640–2653, 2020.
[3] Y. Yang et al., Micromachines, vol. 17, no. 2, p. 184, 2026.
[4] M. Abolghasemian, S. Kaveh, and F. Ebrahimzadeh, Simulation-Based Optimization review, 2025.
[5] S. Liu et al., Complex Intell. Syst., vol. 10, no. 4, pp. 5933–5949, 2024.
[6] Y. Yang et al., IEEE Trans. Circuits Syst. II, vol. 71, no. 4, pp. 2029–2033, 2024.
[7] C. Vişan et al., Knowledge-Based Systems, vol. 258, p. 109987, 2022.
[8] W. Yu et al., Fundamental Research, vol. 4, no. 6, pp. 1442–1454, 2024.
[9] Y. Chen and S. Fang, Electronics Letters, vol. 61, no. 1, p. e70331, 2025.
[10] S. Das, A. Abraham, and A. Konar, Studies in Computational Intelligence, vol. 116, pp. 1–38, 2008.
[11] J. Liu et al., IEEE/ACM ICCAD, pp. 1–9, 2021.
[12] B. Palotai et al., Digital Chemical Engineering, vol. 18, p. 100287, 2026.
[13] M. Sohst, F. Afonso, and A. Suleman, Struct. Multidisc. Optim., vol. 65, no. 1, p. 10, 2022.
[14] R. Gupta and P. H. K. Prasad, Analog Integr. Circ. Sig. Process., vol. 127, no. 2, p. 12, 2026.
[15] N. Shirzadi, D. Lau, and M. Stylianou, Buildings, vol. 15, no. 13, p. 2361, 2025.

Biographies



Umesh Balkishan Phiske received the B.E. degree in Electronics and Telecommunication Engineering in 2011 and the M.E. degree in 2017. He is currently pursuing a Ph.D. at ALARD University, Pune. His research interests include artificial intelligence, machine learning, analog circuit optimization, FPGA-based systems, MATLAB, and NGSpice.



Dr. Gopalkrishna D. Dalvi Guide and HOD, E&TC Engineering, ALARD School of Doctoral Research, ALARD University, Pune – 411057.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

FreshScan: AI-Powered Fruit Quality Assessment System using Image Processing and Deep Learning

Dr. Disha Wankhede[1], Virendrakumar A. Dhotre[2], Mangesh Devikar[3], Siddhant Devsale[4], Neel Dharia[5], Pranit Kothari[6]

Vishwakarma Institute of Technology, Pune, India, disha.wankhede@vit.edu
Nutan Maharashtra Institute of Engineering and Technology, Pune, India, virendrakumar.dhotre@gmail.com
Vishwakarma Institute of Technology, Pune, India, mangesh.devikar231@vit.edu
Vishwakarma Institute of Technology, Pune, India, siddhant.devsale231@vit.edu
Vishwakarma Institute of Technology, Pune, India, neel.dharia23@vit.edu
Vishwakarma Institute of Technology, Pune, India, pranit.kothari23@vit.edu

Abstract.

Poor quality evaluation of fruits is a serious problem with loss of fresh fruit. The visual assessment is subjective and non-predictive. This paper proposes FreshScan, a mobile AI-based real-time fruit quality evaluation and shelf-life prediction system via the deployment of YOLOv8 on the device using TensorFlow Lite. The mean average precision (mAP) is 73-89% for six fruit categories and 300ms for prediction on smartphones. The mean error of the shelf-life prediction is ± 1.2 days and the prediction accuracy is 87.4%. FreshScan is a cost effective and scalable solution which does not need cloud computing resources and addresses this need in making mobile friendly, integrated fruit quality assessment systems.

Keywords. Fruit Quality Assessment; YOLOv8; TensorFlow Lite; Mobile AI; Shelf-Life Prediction

1. INTRODUCTION

Food waste is a worldwide problem and the Food and Agriculture Organization (FAO) states that one-third of the food produced for human consumption is lost and wasted annually [1]. Poor quality assessment at harvest and during sale is a major cause of high loss of fruits from low to high point of sale with a highly perishable nature of the commodity. The current assessment methods are mainly manual, and are based on visual evaluation done by the human operator, leading to lack of uniformity, high error rates and no predictive information regarding shelf life.

Computer vision and deep learning techniques have shown great promise in laboratory environments for fruit quality inspection. We propose FreshScan, an end-to-end, on-device AI solution that (1) detects fruits with the YOLOv8 algorithm fine-tuned for mobile inference using TensorFlow Lite and assess their quality in real-time, and (2) uses a heuristic post-processing model to predict shelf life, embedded within the mobile application.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

A number of computer vision techniques have been discussed recently in the literature for fruit quality detection. CNN based classifiers like VGG-16, ResNet are successful in controlled situations and YOLO based classifiers have started to cater to real-time detection requirements [2,3]. Hyperspectral imaging is capable of analyzing quality at a very fine grain, but requires specialized sensors which are not found in typical consumer devices [4]. Interestingly, there are some works that reach high accuracy, but not a single work that gives the simultaneous real-time deployment of two mobile market and shelf-life prediction as a single system.

3. KEY CONTRIBUTION

The key contributions of this work include real-time, on-device fruit quality detection using YOLOv8 optimized with TensorFlow Lite, achieving approximately 300 ms inference latency on consumer smartphones without cloud dependency. The system also integrates heuristic shelf-life prediction, achieving 87.4% accuracy with a mean error of ±1.2 days. Additionally, it provides a complete end-to-end pipeline from dataset curation and model training to mobile deployment and user interface, with empirical validation across six fruit classes under realistic conditions.

4. METHOD, EXPERIMENTS AND RESULTS

The FreshScan pipeline includes four steps: dataset preparation, model training and optimization, mobile integration and post-processing for quality scoring and shelf-life estimation.

4.1. Dataset Preparation

We compiled a dataset of around 5,000 images of six types of fruits (Apple, Banana, Mango, Orange, Strawberry, and Grape) with labels. Each image was manually annotated with bounding boxes and assigned one of three quality labels: Fresh, Slightly Degraded, or Degraded.

4.2. Model Training

YOLOv8n was selected for its accuracy-inference speed tradeoff. The model was trained 100 epochs. We used a 70/15/15 split for training, validation and test sets, a common choice as it provides enough validation samples to detect early overfitting. This study used data augmentation with random flips, mosaic augmentation and brightness variation.

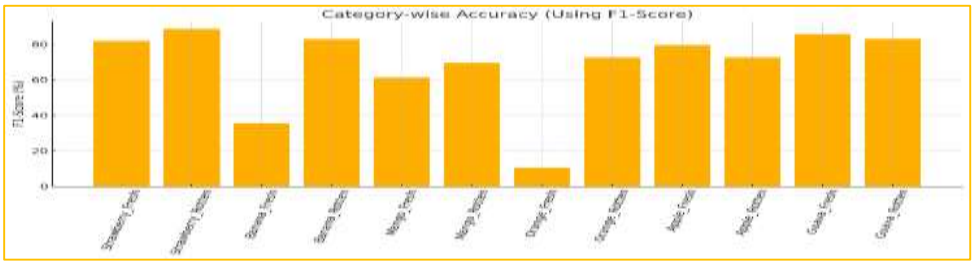


Figure 4.2. Models Performance based on F1-score

4.3. Mobile Integration

The TFLite model is integrated into an Android app. All inference is performed on-device with the TFLite interpreter. Camera frames are captured at 10 FPS and processed in a background thread – this design decision was key to keeping the UI responsive during real-time detection. The final quality score is computed by combining the model's detection confidence with color histogram features. Using both proved more reliable than either alone, particularly under varying lighting conditions.

4.4. Shelf-Life Estimation

A heuristic model is used to produce a map from the quality score, the type of fruit and the environmental temperature (entered by the user or measured by the sensors) to an estimated shelf life of the fruit in days.

Experimental results are presented in Table 2. The system achieves 73–89% mAP across classes, with an average inference time of 297 ms on a mid-range Android device (Snapdragon 665). Shelf-life prediction accuracy is 87.4% with a mean absolute error of ±1.2 days.

Table 1 Class-wise Performance of the Proposed Model for Fruit Detection

Fruit Class	**Precision (%)**	**Recall (%)**	**mAP (%)**
Apple	88.2	85.6	87.1
Banana	91.5	87.3	89.4
Mango	76.8	72.1	74.3
Orange	83.4	79.9	81.6
Strawberry	79.2	74.5	76.8
Grape	71.3	69.8	73.2

5. CONCLUSION

FreshScan offers a convenient mobile technology for evaluating fruit quality and projecting its shelf-life through the new YOLOv8 model and a heuristic post-processing technique. With the use of this technology, the system has demonstrated an accuracy of 73–89% mAP across six fruit classes and 87.4% accuracy in estimating the fruit's shelf-life with the mean error of ±1.2 days. Future efforts are devoted to addressing the problem of class imbalance, enlarging the database, achieving the ability to perform IoT-based environmental monitoring, and extending Fresh Scan's applicability.

6. REFERENCES

[1] FAO, "The State of Food and Agriculture 2019," Food and Agriculture Organization of the United Nations, Rome, 2019.

[2] Y. Jiang, C. Li, "Convolutional neural networks for image-based high-throughput plant phenotyping: A review," Plant Phenomics, vol. 2020, pp. 1–22, 2020. https://doi.org/10.34133/2020/4152816

[3] A. Mazen and A. Nashat, "Ripeness classification of bananas using an artificial neural network," Arabian Journal for Science and Engineering, vol. 44, pp. 6901–6910, 2019. https://doi.org/10.1007/s13369-018-03695-5

[4] A. Nasiri, A. Taheri-Garavand, and Y.-D. Zhang, "Image-based deep learning automated sorting of fresh fruit," Postharvest Biology and Technology, vol. 166, p. 111215, 2020. https://doi.org/10.1016/j.postharvbio.2020.111215

[5] X. Li, C. Rao, Y. Shi, and Q. Liu, "Research on fruit recognition algorithm based on improved YOLOv5," Scientific Programming, vol. 2022, p. 6732500, 2022. https://doi.org/10.1155/2022/6732500

[6] G. Jocher et al., "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[7] TensorFlow Authors, "TensorFlow Lite: On-device ML for Android, iOS, and Edge Devices," 2023. [Online]. Available: https://www.tensorflow.org/li

Biographies



Dr. Disha Sushant Wankhede He is a Postdoctoral Fellow at Universiti Putra Malaysia specializing in AI, Deep Learning, and Healthcare Image Processing. He has 50+ Scopus/WoS publications, 20+ granted patents, and extensive research and industry project experience.



Dr. Virendrakumar A. Dhotre He is an accomplished academician and researcher with 17+ years of academic and nearly 3 years of industry experience. He serves as Assistant Professor and Associate Dean of Academics at VIT Pune, with expertise in IoT, machine learning, data analytics, network security, and intelligent systems.



Mangesh Devikar currently pursuing the bachelor in technology degree in computer engineering from Vishwakarma Institute of Technology, Pune.



Siddhant Devsale currently pursuing the bachelor in technology degree in computer engineering from Vishwakarma Institute of Technology, Pune.



Neel Dharia currently pursuing the bachelor in technology degree in computer engineering from Vishwakarma Institute of Technology, Pune.



Pranit Kothari currently pursuing the bachelor in technology degree in computer engineering from Vishwakarma Institute of Technology, Pune.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Cloud-Based Polyglot Code Execution Framework Using GitHub Integration on Google App Engine

Madhuri D. Kawade[1] , Amit S.Ufade[2]

[1]*Assistant Professor, MVPS'S KBT COE NASHIK*
[2]*Assistant Professor, MET IOE BKC NASHIK*
madhuri3620@gmail.com,amitu_ioe@bkc.met.edu

Abstract.

TCloud computing has become a key platform on which to build and run today's software applications, and distributed software development has been made easy through collaborative version control systems like GitHub. But performing multiple language repositories is still a big challenge as it needs to be configured with compiler, runtime environment, software dependencies, etc., for multiple languages by the developers. This manual process adds up to setup time and leads to variable execution from one system to another. A polyglot code execution system that runs on the cloud system that integrates GitHub and Google App Engine, in order to improve the performance of multi-language software repositories. The proposed framework aims to perform user authentication using the GitHub API, fetch the contents from the user's repository, detect the languages used in the programs and for the project's dependency, and automatically provision the necessary runtime environment in a Platform-as-a-Service (PaaS) infrastructure. Code is executed within the set sandbox environment to ensure security, reliability, and effective use of resources. An experimental analysis indicates that the proposed framework results in a substantial decrease in execution time, increase in scalability and decrease in configuration overhead when compared to the conventional local execution approaches. The framework is a valid approach for cloud software development, DevOps automation, programming labs and enterprise scale application deployment.

Keywords. Google App Engine, GitHub API,polyglot programming, cloud computing, Platform as a Service (PaaS), DevOps, Code Execution Framework.

1. INTRODUCTION

The paradigm of cloud computing has revolutionized software development by offering scalable, flexible, and economical computing resources. Services such as GitHub provide version control, source code management, and collaboration for development projects. Nevertheless, running repositories in multiple programming languages often calls for the configuration of the environment and the setting up of dependencies, which can be time-consuming and impractical. The framework presented here mitigates this issue by dynamically identifying programming languages, choosing appropriate runtimes, and executing code in sandbox environments through the use of PaaS and containerization techniques [5], [7], [10]. The GitHub API will help analyze repositories, and automation enhances security, portability, and scalability [14].

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

2. RELATED WORK

OOne of the key technologies of today's software development and pervasive application deployment is the use of a common computer power, such as servers, storage, on-demand and flexible applications [1]–[3].The characteristics of such resources such as pay-as-you-use, elasticity, scalability utilization etc. allow organizations to manage their fluctuating workloads effectively, and also to reduce the complexity of managing the infrastructure and operational costs [2]. Moreover, By automating the DevOps methods, Continuous Integration/Continuous Delivery, and build, test, and deployment processes.directly from source repositories (CI/CD)pipelines have improved software development efficiency [5], [9]. It is also helpful to be able to deploy distributed and microservice-based applications uniformly across diverse environments with containerization and orchestration platform technologies [5], [7], [10]. But problems like security with multi-language system, seamless runtime provisioning exist still [13]. Thus, there is a need for automated methods to support dynamic runtime provisioning in conjunction with repository analysis, to support scalable and effective execution of polyglot applications in cloud environments.Table 1 shows a comparative analysis of related work in the field of multi-language code execution systems. Table 1: Related Work in Multi-Language Code Execution Systems.

Table 1: Examination of Related Work in Systems for Multi-Language Code Execution.

Ref. no.	Author And Year	Method/ Technology	Key Contribution	Limitation
[2]	Wen et al. (2022)	Serverless Architecture Review	Identifies trends Supporting event-driven Execution suitable code execution frameworks	Limited focus on multi-language (polyglot) execution
[5]	Waseem et al. (2024)	Containerizati on in multi-cloud environments	Enables portable and Scalable deployment Across heterogeneous systems	Requires manual configurati on of runtime environments
[9]	Goli (2024)	DevOps & Cloud Economics	Highlights cost benefits of integrating CI/CD tools like GitHub with cloud platforms.	Lacks technical implementation details
[10]	Ristov et al. (2024)	Federated Serverless (FaaS)	Enables distributed and Scalable execution across cloud platforms	Limited support for Dynamic language detection
[13]	Sun et al.	Software Security	Empirical study on Security patches in	Not directly focused on cloud/serverless

	(2025)	(Multi-language patches)	Open-source systems	computing

3. METHOD, EXPERIMENTS AND RESULTS

The main goals of this proposed research are:1) To create a scalable framework based in the cloud which can run software projects developed in different languages. 2) To create an automated way to get access to source code from GitHub and deploy it using Google's App Engine. To provide an automatic way to identify the programming language of any project and to manage the dependencies automatically without any configured way.3) To ensure a secure way to execute applications from users, within an isolated sandbox.4) To measure the execution time, scalability, resource usage and total performance of the framework against the traditional methods of executing software locally.

3.1. Proposed Method

The new framework will be cloud based, offer product libraries for several programming languages and offer automated compile and run without manual effort. The framework will rely on several components or functional modules, such as user interface, repository management, programming language identification, dependency management, runtime provisioning, secure execution environment and deployment via Google App Engine..

3.2. Experimental Setup and Results Describe scenarios: Execute Performance Tests with: Python, Java and Node.js repositories Measure Execution Latency/Time Perform Scaled Testing (Automatic Scaling) Provide Results in sample performance table (Table 2) that compares and contrasts Execution Setup Times, Scalability, and Resource Usage across different execution methods.

Table 2: Performance Comparison of Manual, Automated (GAE), and Container-Based Execution Approaches.

Setup Time	**High (manual)**	**Low(automated via GAE)**	**Medium(Docker+ Tools)**
Language Support	Limited	Multiple(polyglot)	Multiple
Scalability	Low	High(auto-scaling)	VeryHigh
ResourceCost	Fixed	On-demand	Efficient at scale
CI/CD Integration	Manual	Nativevia GitHub CI	Full DevOps support
Fault Isolation	Low	Medium	High
Runtime Flexibility	Runtime Flexibility	Moderate	High

Table 3: Statistical Validation of Experimental Results

Environment	Mean Execution Time (s)	Std Deviation	Scalability Index
Local Execution	8.84	0.24	Low
Google App Engine	3.3	0.16	High

The statistical findings show that the cloud-based environment has substantially reduced variance and mean execution time.

3.3. Performance Evaluation

The experimental results indicate that the proposed cloud-based architecture can reduce the setup costs and execution time in comparison to the traditional local execution environment. The average of the execution time was reduced from 8.84 seconds in the local environment to 3.30 seconds in the cloud environment, with better consistency and scalability.

4. CONCLUSIONS

In this chapter, a polyglot code execution platform scalable and cloud-based was described, which integrates GitHub and Google App Engine, so as to provide an automated execution of multi-language software repositories. This solution would also eliminate the need to manually configure it, as it would automatically detect programming languages, their dependencies, and provision the appropriate runtime environments. The experimental results show that the proposed solution is more efficient in terms of execution, scalability and reliability when compared with the local way of execution, making it applicable for use in current collaborative software development and DevOps environment.

5. REFERENCES

[1] Z. Li, L. Guo, J. Cheng, Q. Chen, B.He, and M. Guo, “The Serverless Computing Survey: A Technical Primer for Design Architecture,” arXiv:2112.12921, 2021.

[2] J. Wen, Z. Chen, X. Jin, and X. Liu,“Rise of the Planet of Serverless Computing: Systematic Review,” arXiv:2206.12275, 2022.

[3] S. R. Gallardo, “Serverless Strategies And Tools in the Cloud Computing Continuum,” Ph.D. dissertation,Universitat Politècnica de València, 2023.

[4] Q. Zhang et al., “A Survey on LargeLanguage Models for SoftwareEngineering,” arXiv:2312.15223, 2023.

[5] M. Waseem et al., “Containerization in Multi-Cloud Environment: Roles,Strategies, Challenges, and Solutions,”arXiv:2403.12980, 2024.

[6] Thomas, S. Mishra, K. Chen, and J.Kubiatowicz, “Lightweight, Secure and Stateful Serverless Computing with PSL,”arXiv:2410.20004, 2024.

[7] J. Lopes and C. Dbritto, “Implementing Serverless Computing Architectures forExpandable and Cost-Effective Cloud Applications,” International Journal oComputer Trends and Technology, vol. 72,no. 8, pp. 171–179, 2024, doi:10.14445/22312803/IJCTT-V72I8P125.

[8] I.Bansal, “Event-Driven MachineLearning Infrastructure: PerformanceBenchmarking of Cloud Serverless Functions and Container-Based Compute,”International Journal of Intelligent Systems and Applications in Engineering, vol. 12, 2024.

[9] K. R. Goli, “The Economics of Serverless and Containerization: Business Perspectives on DevOps Adoption,” Journal of International Crisis and RiskCommunication Research, 2024, doi:10.63278/jicrcr.vi.3158.

[10] S. Ristov et al., “CODE: Code Once,Deploy Everywhere—ServerlessFunctions in Federated FaaS,” Future Generation Computer Systems, vol. 160,pp. 442–456, 2024, doi:10.1016/j.future.2024.06.017.

[11] C. Lekkala, “Containerization vs. Serverless Architectures for DataPipelines,” SSRN, 2024.

[12] S. Chippagiri, “The Rise of ServerlessComputing: A Systematic Review ofChallenges and Solutions withOptimization Strategies,” Review ofContemporary Scientific and AcademicStudies, vol. 5, no. 1, 2025.:contentReference[oaicite:4]{index=4}

[13] S. Sun, Y. Xing, G. Zou, X. Wang,and K. Sun, “An Empirical Study of Multi-Language Security Patches in OpenSource Software,” in Proc. DIMVA, 2025.

[14] M. K. Alsmadi, A. Al-Marashdeh,and M. Tayfour, “Cloud-NativeTransformation: Architectural Principles and Challenges,” World Journal ofAdvanced Research and Reviews, vol. 26,no. 1, 2025.

[15] “Federated Serverless CloudApproaches: A Comprehensive Review,” Computer and Electrical Engineering, vol.124, 2025.

Biographies



Madhuri D.Kawade received the bachelors degree in computer engineering from Pune University,Pune in 2009, the masters degree in computer engineering from BAMU University in 2014, respectively. She is currently working as an Assistant Professor at the Department of Computer Engineering,Faculty of Engineering,NDMVP’S KBT COE . Her research areas includes Image Processing, deep learning, and IOT.



Dr.Amit S.Ufade received the bachelors degree in Electronics engineering from Pune University,Pune in 2004, the masters degree in E&TC engineering from N.M.U University in 2013, respectively. he is currently working as an Assistant Professor at the Department of Electronics Engineering,MET’S Institute of Engineering BKC Nashik . His research areas include Deep Learning,Human Computer Interface,Automation,image processing.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

A Comprehensive Literature Review on Lean Six Sigma Practices for Sustainable Operational Excellence in the UAE Cement Industry

Mohamad Baraa Ali Sayour[1], Rakesh Kumar Phanden[2], Rajeev Bhushan Garg[3], Vishal Gangadhar Naranje[4]

[1,4] School of Engineering, Architecture and Interior Design, Amity University, Dubai, United Arab Emirates, mohamedB20@amitydubai.ae, vnaranje@amityuniversity.ae

[2]Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India; School of Business and Law, Central Queensland University, 120 Spencer St, Melbourne, VIC, 3000, Australia, rkp.doctor@gmail.com

[3]Star Cement Company LLC (UltraTech Cement), Ras Al Khaimah, UAE, rajeev.g@adityabirla.com

Abstract.

The cement industry is energy and carbon-intensive. Research increasingly integrates sustainability with Lean Six Sigma (LSS), circular economy practices, energy efficiency, and alternative fuels. This review examines success and failure factors, Multicriteria Decision Making (MCDM), DMAIC-based sustainability frameworks, and their validation, providing a basis for a sustainability-driven DMAIC model for the UAE cement industry.

Keywords. Lean Six Sigma (LSS); Sustainable Operational Excellence; Cement Industry; DMAIC; Sustainability.

1. INTRODUCTION

Cement is essential for infrastructure but generates substantial CO2 emissions and industrial waste. Rising energy costs, stricter regulations, and investor expectations have encouraged Lean Manufacturing, Six Sigma, and Green Lean Six Sigma. This review covers sustainability success factors and barriers, MCDM, DMAIC, alternative fuels, and framework validation.

2. EARLY SUSTAINABILITY AND ENERGY EFFICIENCY RESEARCH

Early studies aimed to reduce energy costs and emissions. Following the 1970s fuel crisis, cement manufacturers explored alternative fuels, clinker substitution, waste-heat recovery, and energy-efficient technologies (Coito et al., 2000; Madlool et al., 2011; Hasanbeigi et al., 2011). Later work confirmed alternative fuels, combustion optimization, and clinker substitution as important CO2-reduction routes (Mikulčić et al., 2012; Balsara et al., 2019), establishing a basis for sustainable cement operations.

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

3. LEAN MANUFACTURING AND INTEGRATION WITH LEAN SIX SIGMA

Lean Manufacturing improves productivity by reducing waste and improving process flow (Ruben et al., 2018; Chowdhury et al., 2020). Value Stream Mapping identifies non-value-added activities and constraints (Singh et al., 2017). LSS combines lean waste reduction with Six Sigma's statistical control. Success depends on management commitment, training, communication, and culture (Albliwi et al., 2014; 2015). Green LSS integrates operational, environmental, and economic goals, using energy, greenhouse-gas emissions, and waste as DMAIC indicators (Cherrafi et al., 2016; Hariharan et al., 2021).

4. ENVIRONMENTAL, SOCIAL, AND FINANCIAL SUSTAINABILITY ELEMENTS

Environmental barriers include fossil-fuel dependence, emissions, poor waste management, and regulatory gaps, addressed through emissions control, carbon auditing, clinker substitution, and alternative fuels (Malek and Desai, 2019; Abdullah et al., 2023). Social sustainability relies on leadership, training, communication, and culture, while economic adoption depends on costs, returns, investment feasibility, incentives, and long-term performance (Mangla et al., 2017; McLean et al., 2015; Dube, 2016; Wang, 2016).

5. RENEWABLES, CIRCULAR ECONOMY, AND DECARBONIZATION

Recent literature emphasizes alternative fuels and circular economy integration, including their effects on kiln performance and clinker quality (Larsen, 2007; Saha and Karstensen, 2017; Khokhar et al., 2020). Co-processing, waste recovery, raw-material replacement, supplementary cementitious materials, and pollution prevention are increasingly essential to sustainable cement production (Habert et al., 2020; Hosen and Bărbulescu, 2026).

6. MCDM TECHNIQUES AND SUSTAINABILITY ASSESSMENT

MCDM methods assess sustainability factors and operational barriers. AHP, TOPSIS, VIKOR, and PROMETHEE support manufacturing decisions (Li et al., 2015; Dweiri et al., 2016), while DEMATEL and Fuzzy AHP address causal relationships and uncertainty in circular-economy and green-lean applications (Kashyapa et al., 2022; Garza-Reyes et al., 2025). Integrated MCDM methods provide more robust assessments than standalone techniques.

7. VALIDATION OF SUSTAINABILITY FRAMEWORKS BASED ON DMAIC

DMAIC-based sustainability frameworks are increasingly applied in industry (Mathiyazhagan et al., 2019; Jena et al., 2024). In cement manufacturing, DMAIC has reduced emissions, waste, and energy use, while Statistical Process Control supports stability (Raikar, 2018; Bonetti et al., 2023). LSS frameworks show that DMAIC can integrate operational and environmental improvements (Garza-Reyes et al., 2015; Cherrafi et al., 2016), but empirical validation remains limited, especially in the UAE.

8. RESEARCH GAPS

Most studies treat sustainability and operational excellence separately. The UAE cement sector remains underexplored despite its economic and environmental importance. Few empirical studies combine DMAIC, sustainability KPIs, and alternative fuels to maximize thermal substitution rate (TSR); practical industrial validation is therefore lacking.

9. CONCLUSION

Sustainable cement manufacturing integrates operational, environmental, and economic improvement through LSS, DMAIC, alternative fuels, energy efficiency, circular economy practices, and MCDM. However, the UAE lacks an empirically validated sustainability-DMAIC framework, highlighting the need for future industry-specific research.

10. REFERENCES

[1] Kayikci, S., Khoshgoftaar, T. M., 'Blockchain meets machine learning: A survey', Journal of Big Data, vol. 11, no. 9, 2024. doi: 10.1186/s40537-023-00852-y.

[2] El Hathat, Z., Venkatesh, V. G., Sreedharan, V. R., et al., 'Leveraging greenhouse gas emissions traceability in the groundnut supply chain: Blockchain-enabled off-chain machine learning as a driver of sustainability', Information Systems Frontiers, 2024. doi: 10.1007/s10796-024-10514-w.

[3] Khan, S., Haleem, A., Husain, Z., et al., 'Barriers to blockchain technology adoption in supply chains: The case of India', Operations Management Research, vol. 16, pp. 668–683, 2023. doi: 10.1007/s12063-023-00358-z.

[4] Hirata, E., Lambrou, M., Watanabe, D., 'Blockchain technology in supply chain management: Insights from machine learning algorithms', Maritime Business Review, vol. 6, no. 2, pp. 114–128, 2021. doi: 10.1108/MABR-07-2020-0043.

[5] W. Wisetsri, E. Syam, J. Alanya-Beltran, G. R. Kulkarni, R. K. Vardhan Reddy, M. F. Alam Sheikh, 'Assessing and comparing the role of machine learning (ML) and supply chain management (SCM) towards enhancing e-commerce', Proc. 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, pp. 2574–2577, 2022. doi: 10.1109/ICACITE53722.2022.9823586.

[6] L. Ni, E. Irannezhad, 'Performance analysis of LogisticChain: A blockchain platform for maritime logistics', Computers in Industry, vol. 154, 104038, 2024. doi: 10.1016/j.compind.2023.104038.

[7] K. Duan, G. Pang, Y. Lin, 'Exploring the current status and future opportunities of blockchain technology adoption and application in supply chain management', Journal of Digital Economy, 2024. doi: 10.1016/j.jdec.2024.01.005.

[8] L. Stopfer, A. Kaulen, T. Purfürst, 'Potential of blockchain technology in wood supply chains', Computers and Electronics in Agriculture, vol. 216, 108496, 2024. doi: 10.1016/j.compag.2023.108496.

[9] F. Sunmola, P. Burgess, 'Transparency by design for blockchain-based supply chains', Procedia Computer Science, vol. 217, pp. 1256–1265, 2023. doi: 10.1016/j.procs.2022.12.324.

[10] K. Wannenwetsch, I. Ostermann, R. Priel, F. Gerschner, A. Theissler, 'Blockchain for supply chain management: A literature review and open challenges', Procedia Computer Science, vol. 225, pp. 1312–1321, 2023. doi: 10.1016/j.procs.2023.10.119.

[11] Dong, Z., Liang, W., Liang, Y., et al., 'Blockchained supply chain management based on IoT tracking and machine learning', Journal of Wireless Communications and Networking, 2022, 127, 2022. doi: 10.1186/s13638-022-02209-0.
[12] Peng, X., Zhang, X., Wang, X., et al., 'Construction of rice supply chain supervision model driven by blockchain smart contract', Scientific Reports, vol. 12, 20984, 2022. doi: 10.1038/s41598-022-25559-7.
[13] Biswas, K., Muthukkumarasamy, V., Bai, G., et al., 'A reliable vaccine tracking and monitoring system for health clinics using blockchain', Scientific Reports, vol. 13, 570, 2023. doi: 10.1038/s41598-022-26029-w.
[14] Venkatesan, K., Rahayu, S. B., 'Blockchain security enhancement: An approach towards hybrid consensus algorithms and machine learning techniques', Scientific Reports, vol. 14, 1149, 2024. doi: 10.1038/s41598-024-51578-7.
[15] K. Hande, M. B. Chandak, 'Role of blockchain technology and machine learning in design of smart and secure warehouse management system: A survey', International Journal of Next-Generation Computing, vol. 13, no. 5, pp. 1274–1279, Nov. 2022. doi: 10.47164/ijngc.v13i5.920.
[16] S. Suhail, R. Hussain, R. Jurdak, A. Oracevic, K. Salah, C. S. Hong, R. Matulevičius, 'Blockchain-based digital twins: Research trends, issues, and future challenges', ACM Computing Surveys, vol. 54, no. 11s, Article 240, Jan. 2022. doi: 10.1145/3517189.
[17] Z. Ren, J. Wan, P. Deng, 'Machine-learning-driven digital twin for lifecycle management of complex equipment', IEEE Transactions on Emerging Topics in Computing, vol. 10, no. 1, pp. 9–22, Jan.–Mar. 2022. doi: 10.1109/TETC.2022.3143346.
[18] Ma, X., Du, W., Li, L., et al., 'Examining the nexus of blockchain technology and digital twins: Bibliometric evidence and research trends', Frontiers of Engineering Management, vol. 11, pp. 481–500, 2024. doi: 10.1007/s42524-024-0306-4.
[19] Abrar, I., Sheikh, J. A., 'Current trends of blockchain technology: Architecture, applications, challenges, and opportunities', Discover Internet of Things, vol. 4, no. 7, 2024. doi: 10.1007/s43926-024-00058-5.
[20] B. Unhelkar, S. Joshi, M. Sharma, S. Prakash, A. K. Mani, M. Prasad, 'Enhancing supply chain performance using RFID technology and decision support systems in the Industry 4.0–A systematic literature review', International Journal of Information Management Data Insights, vol. 2, no. 2, 100084, 2022. doi: 10.1016/j.jjimei.2022.100084.
[21] Nazir, J. He, N. Zhu, A. Wajahat, F. Ullah, S. Qureshi, X. Ma, M. S. Pathan, 'Collaborative threat intelligence: Enhancing IoT security through blockchain and machine learning integration', Journal of King Saud University – Computer and Information Sciences, vol. 36, no. 2, 101939, 2024. doi: 10.1016/j.jksuci.2024.101939.
[22] Tufano, R. Accorsi, R. Manzini, 'A machine learning approach for predictive warehouse design', The International Journal of Advanced Manufacturing Technology, vol. 119, no. 3–4, pp. 2369–2392, 2022. doi: 10.1007/s00170-021-08035-w.,

Biographies



Mohamad Baraa Ali Sayour: Mr. Sayour is an academic and industry professional pursuing a Ph.D. in Mechanical Engineering at Amity University Dubai. His expertise includes sustainability, production management, maintenance, and operational efficiency. He has over 10 years of manufacturing industry experience and holds NEBOSH Engineer certification and PMP training.



Dr. Rakesh Kumar Phanden: Dr. Phanden is a senior academic and researcher with over 18 years of international experience. His expertise includes operations and supply chain management, risk management, business analytics, Lean Six Sigma, Industry 4.0, and sustainable operations. He has published over 150 peer-reviewed works and has a Scopus H-index of 20.



Rajeev Bhushan Garg: Mr. Garg is a Cement Technologist with over 33 years of experience in cement manufacturing. He currently leads Production, Quality Control, and Environment at Star Cement Co. LLC, UAE. His expertise includes cement process optimization, plant commissioning, troubleshooting, environmental management, and Six Sigma.



Dr. Vishal G. Naranje: Dr. Naranje is a senior academic and researcher in Mechanical Engineering associated with Amity University Dubai. His research interests include AI in manufacturing, additive and smart manufacturing, digital twins, robotics, industrial engineering, and sustainable engineering. He has published extensively and contributes to academic leadership and industry collaboration.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Streamlining warehouse efficiency: A smart contract approach to blockchain and machine learning integration

Kapil Hande, Manoj Chandak

Shri Ramdeobaba College of Engineering and Management, RTM Nagpur University, Nagpur, kapilhande@gmail.com, chandakmb@rknec.edu

Abstract.

The combination of blockchain technology and machine learning (ML) has the ability to optimize warehouse management systems (WMS) by improving efficiency, transparency, and security. This research, explores blockchain's decentralized, immutable, and transparent ledger, visualizing ML algorithms to improve various aspects of warehouse operations. Blockchain ensures secure and tamper-proof record-keeping, which enhances the reliability of data used for training ML models, while ML provides advanced predictive analytics and real-time decision-making capabilities. The synergy of such technologies addresses key challenges in WMS, including reducing errors, enhancing the accuracy, and automating tasks. This paper proposes a framework for integrating blockchain and ML in WMS, discusses the technical architecture, and evaluates the performance improvements in terms of operational efficiency and security. The application of advanced Long-Short Term Memory (LSTM) networks on data generated using Amazon sales data processed using Probability Density grouping achieved a Pearson Correlation of 0.9908 and RMSE of 0.0200. Our findings demonstrate that the hybrid approach not only streamlines warehouse processes but also fosters trust and collaboration across sup-ply chain stakeholders.

Keywords. Blockchain Technology; Machine Learning; Warehouse Management Systems; Smart Contracts; Supply Chain Management.

1. INTRODUCTION

Supply chain management controls the management of products and services from raw material sourcing to final product handover. It consists of managing the movement, storage, and inventory at each stage of production, as well as coordinating suppliers to make sure smooth transitions from source to destination. This process includes handling work-in-progress and completed goods, forming a network of suppliers to facilitate efficient order fulfillment. Warehouses are crucial to the supply chain, providing not only as storage facilities but also as key points for shipping and receiving goods. As warehouse operations grow more complex, the use of technology, particularly machine learning algorithms, is becoming essential for improving efficiency, re-ducing costs, and boosting productivity.[1][2][3]

Machine learning enables the discovery of pat-terns in the supply chain by using algorithms to ana-lyze performance and identify areas for improvement more efficiently than manual methods. It significant-ly impacts inventory levels, quality, supply and de-mand, production planning, and transport manage-ment, especially in warehouse management.[4]

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

Blockchain technology improves transparency and accuracy in end-to-end supply chain tracking by digitizing physical assets and providing a decentral-ized, immutable record of transactions. This allows organizations to track assets from production to final delivery or use, providing greater visibility for both businesses and consumers. For high-value items like diamonds and pharmaceutical drugs, blockchain re-duces fraud by enabling companies to trace the flow of ingredients and products through subcontractors, minimizing losses from counterfeit and gray market activities while boosting consumer confidence in the authenticity of goods. Blockchain, as a decentralized distributed ledger, abolish the need for a central au-thority or administrator for managing data records and transactions.[5][6] Every block in the blockchain contains a cryptographic hash linking it to the earlier block, ensuring the immutability of the entire chain.

When integrated with the Internet of Things (IoT), blockchain can enhance logistics by enabling faster deliveries and reducing operational costs across various applications.[7][8] By introducing au-tonomy and transparency into logistics processes, blockchain improves both security and operational efficiency.[9][10]

Emerging technologies like Distributed Ledger Technology (DLT) and IoT offer solutions for ad-dressing challenges related to tracking and prove-nance in supply chains.

Predictive data analytics employs current and histor-ical data to forecast future events or unknown out-comes. By analyzing this data, it allows for the as-sessment of the likelihood of specific situations or occurrences.[11-13] Unlike general forecasting, pre-dictive data analytics provides insights at the indi-vidual level—pertaining to specific customers, em-ployees, or systems.

Smart contracts are self-executing, immutable, and distributed code deployed on the blockchain, written in Solidity.[14][15] They are tamper-proof and validated by everyone on the network, ensuring secure and transparent transactions. Once deployed, smart contracts cannot be altered, allowing them to govern transactions and decision-making processes. These contracts enable the transparent transfer of as-sets—such as money, shares, or property—without intermediaries. Like traditional contracts, they define rules and penalties, but smart contracts automatically enforce the agreed conditions.

Remaining paper is organized as follows. Section II discusses the ongoing research scenario of the supply chain. Section III throws light on the various processes involved in the development of smart WMS. Section IV demonstrates the implementation part along with dataset used and results developed. Finally Section V concludes the paper.

demands without considering the data volume. Taking into consideration this ratio, green IT technologies have important benefits in terms of:

2. RELATED WORK

Z. Ren et. al. [17] explores how integrating digital twin technology with machine learning enhances the full life cycle management of complex equipment, crucial for modern manufacturing's intelligent trans-formation. It proposes a system embedding machine learning into digital twins and constructs a life cycle digital twin for complex equipment. Experimental results show that abnormal axle temperatures can be detected a week in advance, demonstrating the method's effectiveness.

Safak Kayikci et. al. [1] reviews the contributions of machine learning and blockchain integration, in-cluding enhanced security, privacy, and decentraliza-tion. Blockchain ensures reliable, tamper-proof de-centralized record-keeping, whereas machine learn-ing provides the capability to extract meaningful in-sights from large-scale data. Together, these technol-ogies promise to reshape industries by streamlining operations, fostering intelligent and secure systems, empowering informed decision-making, and fortify-ing security by mitigating vulnerabilities and safe-guarding data integrity.

Venkatesan, K et. al. [14] introduces hybrid con-sensus algorithms enhanced by machine learning (ML) approaches to manage the limitations and vul-nerabilities of blockchain networks. While traditional consensus protocols often struggle with assuring compliance and are prone to cyber-attacks, this re-search proposes novel approaches like Delegated Proof of Stake Work (DPoSW). By integrating ML for anomaly detection, cyber-attack prediction, and feature extraction, the proposed framework enhances the adaptability and energy efficiency of blockchain systems. However, challenges such as scalability, la-tency, resource demands, and adversarial threats re-main key hurdles for real-world implementation.

Claudia Antal et. al. examines how blockchain technology can enhance the vaccination process. It proposes a system that ensures data integrity and prevents identity theft through secure beneficiary registration. Smart contracts are used to monitor vaccine distribution in accordance with producers' safe handling guidelines, keeping all network partic-ipants informed. A prototype was developed on the Ethereum test network, Ropsten, with results availa-ble for verification on Etherscan.

Ahsan Nazir et. al. [21] introduces a new method to improve IoT security by combining coordinated threat intelligence with blockchain technology and machine learning (ML) models. An iOS application serves as a central hub for reporting and sharing de-tected threats, with data securely stored on a block-chain network. The study employs Random Forest, Decision Tree, Ensemble, LSTM, and CNN models on the IoT23 dataset within Mutual Threat Intelli-gence architecture.

Li Li et. al. introduces a secure traceability framework integrating blockchain technology, a cen-tralized database, and RFID tags. By leveraging blockchain's immutability for summary data and op-timizing the SM3 algorithm, the system achieves a 30% improvement in efficiency, reducing execution time for 192-byte messages to 210μs. Traceability data is securely time-stamped and linked, with de-tailed information retained in the centralized data-base while unalterable summaries are recorded on the blockchain.

3. BACKGROUND

Warehouse management system as a part of warehouse is becoming complex task because of its operations and volume of data it handle. The purpose of this research is to demonstrate the employ blockchain technology for secure transactions and machine leaning as predictive data analytics tool. The goal is to improve the efficiency of operations of warehouse and optimize the performance parameters. The transaction level information can be stored in the blockchain and smart contracts can be used to carry out the transactions processing. This

confirms the transparency between the parties involved in the business. Smart contracts play a role of online agreements based on meeting certain conditions.

The major issue in the warehouse is unpredictability of the operations required to fulfill the order and item request.[16] Product handling processes in the warehouse generates the large amount of data. This data needs to be recorded, stored and analyzed for predictive analysis. The data generated will be stored in the blocks of the blockchain server. Hence the physical space data is converted into the digital space for predictive data analysis through machine learning. Barcode system can be incorporated for digital tagging of the warehouse items and processes.[17-18]

Figure 1 illustrates the key components in designing a smart warehouse management system, highlighting the transformative roles of blockchain technology and machine learning. These cutting-edge technologies have applications across various fields, including supply chain management, healthcare, and manufacturing. In warehouse management systems (WMS), critical challenges such as space and route optimization, rapid item retrieval, demand forecasting, and transaction management are addressed. Smart contracts provide enhanced security, operational transparency, automated payments, and streamlined inventory management, while the machine learning module drives efficiency across the entire system.[19-20]

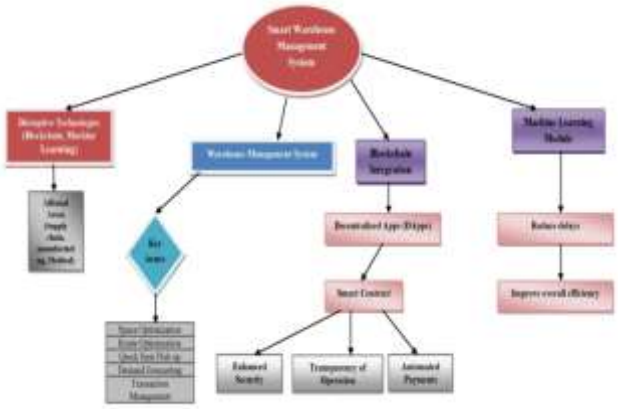


Figure 3.1: Aspects of Smart WMS

4. METHODOLOGY

4.1. Steps in consideration in WMS

In warehouse management system the process flow is as follows. Order Acceptance is a process of confirming and accepting an order received from a customer. Inventory Check is the procedure to verify the availability of items in stock before processing the order. Picking is the action of retrieving the items from the warehouse shelves as per the order requirements. Packing is the process of securely packing the picked items for shipment.

Documentation phase prepare and organize necessary paperwork and labels for the shipment. Carrier Coordination step arrange and schedule with the carrier service for order

pickup and delivery. Wait for Carrier is the time during which the packed orders are waiting for the carrier to arrive for pickup. Loading is the process of loading the packed orders onto the carrier's vehicle for transport.[21][22]

4.2. Generation of Time Data

We have used the Amazon sales report dataset for data analytics and model building. The contents of the data set are depicted in table 1.

Table 1. Amazon Sales Report

Name of the Data Set	Amazon Sale Report
No of records	128975
Item timing parameters generated	Wait for carrier
	Order acceptance adjusted
	Inventory check adjusted
	Picking adjusted
	Packing adjusted
	Documentation adjusted
	Carrier coordination adjusted
	Loading adjusted
	Total Time

Following functions are used to generate the synthetic time data.

- **Time Ranges Based on Buying Patterns**: The function defines four specific time ranges representing typical buying patterns throughout the day: morning peak (8 AM to 12 PM), afternoon peak (12 PM to 3 PM), evening peak (5 PM to 8 PM), and off-peak (8 PM to 12 AM).
- **Weighted Selection of Time Ranges**: Each time range is assigned a weight reflecting the likelihood of purchases occurring within that period. These weights simulate observed buying behavior: morning peak (30%), afternoon peak (40%), evening peak (25%), and off-peak (5%).

- **Random Selection within Time Ranges**: Using the weights, a time range is randomly selected to simulate the time of purchase. This selection process ensures that the synthetic time data aligns with typical buying patterns.
- **Random Time Generation within Selected Range**: After selecting a time range, the function generates a random time within this range by randomly selecting an hour, minute, and second. This approach produces a diverse set of times while adhering to the overall buying behavior trends.

Pseudo Code for generation of time data

```
Function generate_synthetic_time():
    Define time ranges:
morning_peak = (8, 12)        // 8 AM to 12 PM
afternoon_peak = (12, 15)     // 12 PM to 3 PM
evening_peak = (17, 20)       // 5 PM to 8 PM
off_peak = (20, 24)           // 8 PM to 12 AM

    Define weights for each time range:
weights = [0.3, 0.4, 0.25, 0.05]

    Select a time range based on weights:
selected_range = random.choices(time_ranges, weights)

    Generate a random time within the selected range:
hour = random.randint(selected_range[0], selected_range[1] - 1)
minute = random.randint(0, 59)
second = random.randint(0, 59)
    Return time(hour, minute, second)
```

4.3. Generation of Raw Elapse Time data for each step of WMS

For each step in the WMS process (e.g., Order Acceptance, Inventory Check), a base time (in minutes) and an associated randomness (standard deviation) were defined. These values represent the average expected time and the variability for each step. To generate realistic time data that reflects the natural variability of each step while ensuring all values are positive and non-zero, a truncated normal distribution was used. The function uses PDF (Probability Density Function) and CDF (Cumulative Distribution Function). This

distribution is bounded to ensure times remain within a plausible range, avoiding negative or excessively high values. Following is the formula for truncated normal distribution.

$$f(x|\mu,\sigma,a,b) = \frac{\phi(\frac{x-\mu}{\sigma})}{\sigma((\frac{b-\mu}{\sigma})-\phi(\frac{a-\mu}{\sigma}))} \quad (4.3.1)$$

$\phi(\frac{x-\mu}{\sigma})$ is the PDF of standard normal distribution

$\phi(\frac{a-\mu}{\sigma})$ is the CDF of standard normal distribution

For each step, the mean and standard deviation were set to the base time and randomness, respectively. The bounds were determined to be slightly above zero (0.01) and up to three times the base time, ensuring realistic yet varied durations. For each step, the truncated normal distribution function was applied to generate a column of times. This function ensured that the generated times varied around the base time according to the defined randomness, producing realistic and variable step durations for each entry in the dataset.

4.4. Adjustment for load in elapsed time of WMS steps

The function for a load calculation calculates the current load by counting the number of records in the dataset that have the same hour as the current row's timestamp. This load represents the number of tasks or processes occurring within that specific hour. The raw elapsed time for a given step is adjusted using a load factor, which scales the time based on the calculated load. The formula used is

adjusted time=raw time x (1 + load factor x load)

This adjustment increases the elapsed time proportionally to the load, simulating delays or extended processing times due to higher load. A lambda function is used to calculate and apply adjustment to time data of each step of WMS. This lambda function is passes into df.apply function of dataframe which applies the adjustment to each step according to calculated load. A simple arithmetic sum is taken for all the adjusted times of WMS steps to calculate total time taken for all steps to complete within the WMS. This field is added to the dataframe for further usage. The table 2 below shows the statistics of the elapsed time for given dataset.

Table 2. Statistics of Adjusted Times

Name	Min (min.)	Max (min.)	Mean (min.)	Mean (hr.)	Mean (days)
Order Acceptance	0.14	1574.17	417.47	6.96	0.29
Inventory Check	0.15	6531.87	1670.18	27.84	1.16
Picking	0.22	1294.98	406.87	6.78	0.28
Packing	0.72	310.22	108.42	1.81	0.08

Documentation	0.08	528.57	139.03	2.32	0.10
Carrier Coordination	0.20	267.75	94.83	1.58	0.07
Loading	0.45	398.50	135.42	2.26	0.09
Wait for Carrier	0.47	8619.00	2960.71	49.35	2.06
Total Time	137.32	9084.79	2972.23	49.54	2.06

4.5. Demand Prediction

In data preprocessing phase the dataset is read and preprocessed by converting datetime columns, handling missing values, and encoding categorical variables. Numerical features are scaled using StandardScaler to ensure the data is centered around zero with a unit standard deviation, thereby enhancing the effectiveness of the machine learning model. The time series data is prepared using TimeseriesGenerator, which creates sequences of a specified length (sequence_length) from the scaled features to be used as input for the LSTM model. This generator pairs the input sequences with the corresponding target values (quantities). An LSTM model is built using Keras, consisting of an LSTM layer with 50 units and a Dropout layer to prevent overfitting. The model is trained on the generated sequences to predict the quantity (Qty) of orders. The training is conducted for a predefined epochs, during which the model learns to capture the temporal dependencies in the data.

After training, the model's predictions are evaluated against the actual values using metrics. The model showed Pearson correlation coefficient of 0.9908 showing very accurate demand prediction. The model achieved a very low MSE loss of 0.0165 and RMSE of 0.0200. The actual and predicted quantities are plotted over different time intervals (daily, hourly, weekly) to visualize the model's effectiveness and the reliability of demand predictions. The following figure 2 shows the graph of actual versus predicted daily quantity over time.

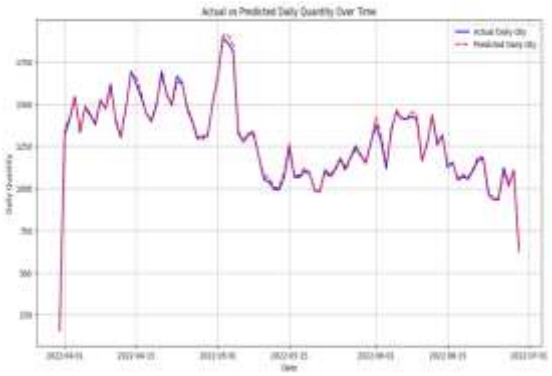


Figure 4.5.1. Plot of Actual vs Daily Quantity over Time

5. CONCLUSION

In this research, we analyzed the integration of Blockchain technology and machine learning to develop a smart Warehouse Management System (WMS). By combining these two transformative technologies, we aimed to improve the performance of traditional WMS. Using an Amazon sales report as the dataset, we generated synthetic timestamps to simulate real-time conditions and analyzed each WMS material handling step accordingly. An LSTM model was constructed to evaluate the data, incorporating a demand prediction module to track actual versus forecasted daily quantities over time, alongside a total turnaround time prediction. Our results demonstrate improved accuracy in demand forecasting and reduced turnaround times within the experimental data model, underscoring the potential of our approach to optimize WMS performance.

6. REFERENCES

[1] Kayikci, S., Khoshgoftaar, T.M., 'Blockchain meets machine learning: a survey', *Journal of Big Data*, vol. 11, no. 9, 2024. doi: 10.1186/s40537-023-00852-y.

[2] El Hathat, Z., Venkatesh, V.G., Sreedharan, V.R., *et al.*, 'Leveraging greenhouse gas emissions traceability in the groundnut supply chain: Blockchain-enabled off-chain machine learning as a driver of sustainability', *Information Systems Frontiers*, 2024. doi: 10.1007/s10796-024-10514-w.

[3] Khan, S., Haleem, A., Husain, Z., *et al.*, 'Barriers to blockchain technology adoption in supply chains: The case of India', *Operations Management Research*, vol. 16, pp. 668–683, 2023. doi: 10.1007/s12063-023-00358-z.

[4] Hirata, E., Lambrou, M., Watanabe, D., 'Blockchain technology in supply chain management: Insights from machine learning algorithms', *Maritime Business Review*, vol. 6, no. 2, pp. 114–128, 2021. doi: 10.1108/MABR-07-2020-0043.

[5] Wisetsri, W., Syam, E., Alanya-Beltran, J., Kulkarni, G. R., Vardhan Reddy, R. K., Alam Sheikh, M. F., 'Assessing and comparing the role of machine learning (ML) and supply chain management (SCM) towards enhancing e-commerce', *Proc. 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, pp. 2574–2577, 2022. doi: 10.1109/ICACITE53722.2022.9823586.

[6] Ni, L., Irannezhad, E., 'Performance analysis of LogisticChain: A blockchain platform for maritime logistics', *Computers in Industry*, vol. 154, 104038, 2024. doi: 10.1016/j.compind.2023.104038.

[7] Duan, K., Pang, G., Lin, Y., 'Exploring the current status and future opportunities of blockchain technology adoption and application in supply chain management', *Journal of Digital Economy*, 2024. doi: 10.1016/j.jdec.2024.01.005.

[8] Stopfer, L., Kaulen, A., Purfürst, T., 'Potential of blockchain technology in wood supply chains', *Computers and Electronics in Agriculture*, vol. 216, 108496, 2024. doi: 10.1016/j.compag.2023.108496.

[9] Sunmola, F., Burgess, P., 'Transparency by design for blockchain-based supply chains', *Procedia Computer Science*, vol. 217, pp. 1256–1265, 2023. doi: 10.1016/j.procs.2022.12.324.

[10] Wannenwetsch, K., Ostermann, I., Priel, R., Gerschner, F., Theissler, A., 'Blockchain for supply chain management: A literature review and open challenges', *Procedia Computer Science*, vol. 225, pp. 1312–1321, 2023. doi: 10.1016/j.procs.2023.10.119.

[11] Dong, Z., Liang, W., Liang, Y., *et al.*, 'Blockchained supply chain management based on IoT tracking and machine learning', *Journal of Wireless Communications and Networking*, 2022, 127, 2022. doi: 10.1186/s13638-022-02209-0.
[12] Peng, X., Zhang, X., Wang, X., *et al.*, 'Construction of rice supply chain supervision model driven by blockchain smart contract', *Scientific Reports*, vol. 12, 20984, 2022. doi: 10.1038/s41598-022-25559-7.
[13] Biswas, K., Muthukkumarasamy, V., Bai, G., *et al.*, 'A reliable vaccine tracking and monitoring system for health clinics using blockchain', *Scientific Reports*, vol. 13, 570, 2023. doi: 10.1038/s41598-022-26029-w.
[14] Venkatesan, K., Rahayu, S. B., 'Blockchain security enhancement: An approach towards hybrid consensus algorithms and machine learning techniques', *Scientific Reports*, vol. 14, 1149, 2024. doi: 10.1038/s41598-024-51578-7.
[15] Hande, K., Chandak, M. B., 'Role of blockchain technology and machine learning in design of smart and secure warehouse management system: A survey', *International Journal of Next-Generation Computing*, vol. 13, no. 5, pp. 1274–1279, Nov. 2022. doi: 10.47164/ijngc.v13i5.920.
[16] Suhail, S., Hussain, R., Jurdak, R., Oracevic, A., Salah, K., Hong, C. S., Matulevičius, R., 'Blockchain-based digital twins: Research trends, issues, and future challenges', *ACM Computing Surveys*, vol. 54, no. 11s, Article 240, Jan. 2022. doi: 10.1145/3517189.
[17] Ren, Z., Wan, J., Deng, P., 'Machine-learning-driven digital twin for lifecycle management of complex equipment', *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 1, pp. 9–22, Jan.–Mar. 2022. doi: 10.1109/TETC.2022.3143346.
[18] Ma, X., Du, W., Li, L., *et al.*, 'Examining the nexus of blockchain technology and digital twins: Bibliometric evidence and research trends', *Frontiers of Engineering Management*, vol. 11, pp. 481–500, 2024. doi: 10.1007/s42524-024-0306-4.
[19] Abrar, I., Sheikh, J. A., 'Current trends of blockchain technology: Architecture, applications, challenges, and opportunities', *Discover Internet of Things*, vol. 4, no. 7, 2024. doi: 10.1007/s43926-024-00058-5.
[20] Unhelkar, B., Joshi, S., Sharma, M., Prakash, S., Mani, A. K., Prasad, M., 'Enhancing supply chain performance using RFID technology and decision support systems in the Industry 4.0–A systematic literature review', *International Journal of Information Management Data Insights*, vol. 2, no. 2, 100084, 2022. doi: 10.1016/j.jjimei.2022.100084.
[21] Nazir, A., He, J., Zhu, N., Wajahat, A., Ullah, F., Qureshi, S., Ma, X., Pathan, M. S., 'Collaborative threat intelligence: Enhancing IoT security through blockchain and machine learning integration', *Journal of King Saud University – Computer and Information Sciences*, vol. 36, no. 2, 101939, 2024. doi: 10.1016/j.jksuci.2024.101939.
[22] Tufano, A., Accorsi, R., Manzini, R., 'A machine learning approach for predictive warehouse design', *The International Journal of Advanced Manufacturing Technology*, vol. 119, no. 3–4, pp. 2369–2392, 2022. doi: 10.1007/s00170-021-08035-w.

Biographies



Mr Kapil Hande is research scholar at Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur. He has total 26 years of teaching experience. His research areas includes Cryptography, Artificial Intelligence, Machine learning and Blockchain technology. He has completed MODROB and DST funded projects and also guided several PG projects. He has published a book on Operating System. He can be contacted at email: kapilhande@gmail.com.



PROF. Manoj Chandak is Principal at Shri Ramdeobaba College of Engineering and Management, Nagpur. India. He has total of 27 years of academic experience. His areas of interest are Natural Language Processing, Machine Learning, Big Data Analytics, Cloud storage, and data management. His current research work includes the "Development of coal quality exploration technique based on convolutional neural networks and hyper spectral images". The project received a grant of INR 1.03 cr . Prof. Chandak is also the principal investigator of the project which received a grant of 1.10Cr from the Ministry of Coal, Government of India. He can be contacted at email: chandakmb@rknec.edu

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

An Enhanced 3D U-NET CNN for Brain Segmentation and Grading

Dr. Y. Ambica[1], M. Abhiram Seshu[2], J. Nandini[3]

[1-3]CSM, CMR College of Engineering & Technology, Medchal, Telangana, India
dr.y.ambica@gmail.com[1] , abhiramseshu@gmail.com[2] , jayanthinandini448@gmail.com[3]

Abstract.

Magnetic Resonance Imaging (MRI) is widely used for brain tumor diagnosis, but manually analyzing large volumes of MRI data is time-consuming. This paper proposes an automated machine learning pipeline for brain tumor segmentation, grading, and volume measurement using T1, T1ce, T2, and FLAIR MRI images from the BraTS 2020 dataset. A MONAI-implemented U-Net segments Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), while a rule-based module performs tumor grading and voxel-based volume estimation. The system achieved a Dice score of 0.77 for WT, while TC and ET achieved 0.10 due to class imbalance and complex tumor structures. Despite these challenges, the unified pipeline can reduce analysis time and support radiologists in diagnosis, treatment planning, and disease monitoring.

Keywords: Brain Tumor Segmentation, MRI, 3D U-Net, Tumor Grading, Volumetric Analysis.

1. INTRODUCTION

Brain tumors require rapid and accurate diagnosis for effective treatment and disease monitoring. MRI provides detailed soft-tissue information through T1, T1ce, T2, and FLAIR sequences, but manual interpretation is time-consuming and challenging due to tumor variability and differences in imaging protocols. Deep learning, particularly 3D U-Net, improves tumor segmentation by capturing complete 3D spatial information. While segmentation identifies tumor regions, grading assesses tumor aggressiveness, and volumetric analysis measures tumor size. Therefore, this work proposes an integrated system using 3D U-Net for segmentation, rule-based grading, and volume estimation from multimodal MRI data.

2. RELATED WORK

Brain tumor detection and segmentation have evolved from traditional image-processing methods to machine learning and deep learning [3]. Early approaches used MRI segmentation and texture analysis , [5], while diffusion imaging and self-organizing maps were applied for tumor grading. Multimodal MRI methods improved tumor boundary detection by combining spatial and intensity features [2], with studies showing better performance than single-modality scans [1] and reduced computational cost [8]. Hybrid and machine learning approaches further improved automatic tumor detection and segmentation. Recent deep learning methods include CNNs with transfer

learning [6], 2D V-Net, SegNet, and modified U-Net architectures. Attention and transformer-based models such as TransUNet and pyramid convolution with attention gates, along with multitask learning [7], have further advanced glioma segmentation. Reviews have summarized these developments and highlighted the widespread use of U-Net-based approaches [4], while integrated diagnostic systems combine segmentation with classification.

Table 1: Comparative Analysis of Existing Methods and Proposed System

Category	Method / Model	Strengths	Limitations
Traditional Methods	Thresholding, Region Growing, Edge Detection	Low computational cost, simple implementation	Highly sensitive to noise, poor accuracy for complex tumors
Classical Machine Learning	SVM, KNN, Random Forest	Better than rule-based methods, interpretable	Feature engineering required, limited generalization
Proposed Approach	3D U-Net (MONAI + PyTorch)	Accurate tumor delineation, clinical relevance	Training time, hardware dependency

2.1. Key Contribution

The main contribution of this work is that it builds an integrated framework that performs outlining, grading, and size measurement of brain tumors in one automated system. A 3D U-Net built on the MONAI platform is used to outline tumors from multimodal MRI scans.

3. METHODOLOGY

This methodology provides an automatic process for multimodal MRI brain tumor analysis. The system. Measures tumors by combining outlining, grading and size measurement into one flow. A 3D convolutional neural network trained with learning looks at the whole MRI volume and gives useful clinical data. The BraTS 2020 dataset supplies multimodal MRI scans, including T1, T1 contrast-enhanced (T1ce) T2 and FLAIR. Because MRI images vary in resolution and angle**,** a preprocessing stage is needed.

Table 3: Network Architecture Configuration

Parameter	Value
Optimizer	Adam
Learning Rate	1e-4
Loss Function	Dice Loss
Batch Size	1
Number of Epochs	5

Device	CUDA / CPU
Patch Size	128 × 128 × 128
Data Augmentation	Random Crop, Flip, Rotation
Label Mapping	Class 4 → 3

Table 4: Network Architecture Configuration

Parameter	**Value**
Model Type	3D U-Net (MONAI)
Input Channels	4 (T1, T1c, T2, FLAIR)
Output Channels	4 (Background + 3 Tumor Classes)
Spatial Dimensions	3D
Encoder Channels	(16, 32, 64, 128, 256)
Number of Levels	5
Strides	(2, 2, 2, 2)
Number of Residual Units	2 per level
Kernel Size	3 × 3 × 3
Activation Function	PReLU
Normalization	Instance Normalization
Final Activation	Softmax

4. RESULT ANALYSIS

The proposed brain tumor analysis system was evaluated on the BraTS 2020 dataset for segmentation, grading, and volumetric analysis. The 3D U-Net effectively extracted spatial features from multimodal MRI images, particularly FLAIR and T1ce. The model achieved a Dice score of 0.77 for Whole Tumor (WT), while Tumor Core (TC) and Enhancing Tumor (ET) showed lower performance due to class imbalance and smaller tumor regions. High specificity across classes indicates effective reduction of false-positive predictions.

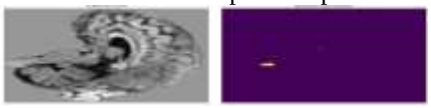


Figure 3. 2 Tumor Segmentation of 3D MRI

Table 4 : Performance Metrics

	Dice	**Sensitivity**	**Specificity**
WT	0.77	0.70	0.998
TC	0.20	0.39	0.989

ET	0.05	0.53	0.994

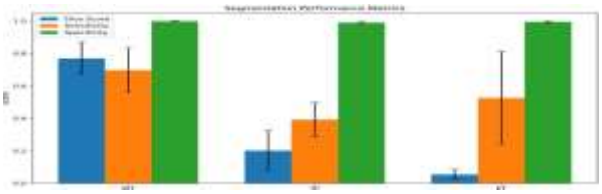


Figure 3.3: Segmentation Performance metrics

5. DISCUSSION

The proposed system integrates deep learning for brain tumor segmentation, grading, and volumetric analysis, providing pixel-level interpretation through a decoupled pipeline. It performs well in Whole Tumor segmentation but struggles with smaller regions such as Tumor Core and Enhancing Tumor due to class imbalance. Future improvements could include attention and transformer-based architectures. Despite computational costs and lower accuracy for small tumor regions, the framework demonstrates useful clinical potential.

6. CONCLUSION

The proposed system integrates MRI preprocessing, 3D U-Net-based tumor segmentation, rule-based grading, and volume estimation into a single automated pipeline. Using multimodal MRI data, the model achieved a Dice score of 0.77 for Whole Tumor segmentation. Its modular and interpretable design provides clinically useful information for diagnosis, treatment planning, and efficient clinical decision-making.

7. REFERENCES

[1] Safia Fatima, Hafeez Ur Rehman, Aizaz Sharif, Usman Habib, "Evaluation of Multi-Modal MRI Images for Brain Tumor Segmentation",15th International Conference on Emerging Technologies (ICET),IEEE,2019, DOI: 10.1109/ICET48972.2019.8994408.

[2] Syed Muhammad Anwar, Sobia Yousaf, Muhammad Majid, "Brain Tumor Segmentation on Multimodal MRI Scans Using EMAP Algorithm", 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE,2018 DOI: 10.1109/EMBC.2018.8512304.

[3] Hrishikesh Lamdade, Arjun Pansare, Gaurav Parulekar, Jignesh Sisodia, "Brain Tumor Segmentation, Grade of Tumor and Survival Duration Prediction Using Deep Learning", 10th International Conference on Signal Processing and Integrated Networks(SPIN), IEEE,2023 DOI: 10.1109/SPIN57001.2023.10117418.

[4] Mina Ghaffari, Arcot Sowmya, Ruth Oliver, "Automated Brain Tumor Segmentation Using Multimodal Brain Scans: A Survey Based on Models Submitted to the BraTS

2012–2018 Challenges", IEEE Reviews in Biomedical Engineering, Vol. 13, IEEE, 2020, ISSN No: 1937-3333 DOI: 10.1109/RBME.2019.2946868.

[5] Atiq Islam, Syed M. S. Reza, Khan M. Iftekharuddin, "Multifractal Texture Estimation for Detection and Segmentation of Brain Tumors", IEEE Transactions on Biomedical Engineering, Vol. 60, Issue 11, IEEE, 2013, ISSN No: 0018-9294 DOI: 10.1109/TBME.2013.2271383.

[6] Sinan Alkassar, Mohammed A. M. Abdullah, Bilal A. Jebur, "Automatic Brain Tumour Segmentation Using Fully Convolution Network and Transfer Learning", 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE), IEEE, 2019 DOI: 10.1109/ICECCPCE46549.2019.203771.

[7] Sakshi Ahuja, B. K. Panigrahi, Tapan Gandhi, "Transfer Learning Based Brain Tumor Detection and Segmentation Using Superpixel Technique", International Conference on Contemporary Computing and Applications (IC3A), IEEE, 2020 DOI: 10.1109/IC3A48958.2020.233306.

[8] Fateme Mostafaie, Reihaneh Teimouri, Zahra Nabizadeh Shahre Babak, Nader Karimi, Shadrokh Samavi, "Region of Interest Identification for Brain Tumors in Magnetic Resonance Images", 28th Iranian Conference on Electrical Engineering (ICEE), IEEE, 2020 DOI: 10.1109/ICEE50131.2020.9260611.

Biographies



Dr.Y Ambica received the bachelor's degree in computer engineering from JNTUH University in 2005, the master's degree in computer engineering from JNTUH University in 2010, and the philosophy of doctorate degree in Computer Science Engineering from JNTUH University in 2023, respectively.



Jayanthi Nandini received her Bachelor's degree in Computer Science and Engineering, specializing in Artificial Intelligence and Machine Learning, from CMR College of Engineering and Technology (CMRCET).



Mathukumalli Abhiram Seshu received his Bachelor's degree in Computer Science and Engineering, specializing in Artificial Intelligence and Machine Learning, from CMR College of Engineering and Technology (CMRCET).

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

WeSee – A Helper for the Visually Impaired

Rutik Tupe[1], Pooja Thorat[2], Sarthak Velapure[3], Prof . Shraddha Toney[4]

[1,2,3,4]Department of Computer Engineering, AISSMS Institute of Information Technology, Pune, India

[1]rutikrt1408@gmail.com , [2]poojathorat1410@gmail.com ,[3]Sarthakvelapure5644@gmail.com ,[4]Shraddha.toney@aissmsioit.org

Abstract

Navigating everyday environments and identifying surrounding objects is a major challenge for people with visual impairments. Traditional aids such as the white cane support basic obstacle avoidance but do not provide contextual information about nearby objects. This paper presents WeSee, an AI-driven assistive web application that improves situational awareness for the visually impaired. Built on a Flask framework, WeSee captures real-time environmental data and applies a YOLO-based computer-vision model for instant object recognition, converting the results into spoken feedback through a text-to-speech engine. By combining microservice-style AI modules with an SQLite backend, WeSee helps users navigate independently. Testing confirms high precision in identifying common indoor objects while maintaining low-latency audio responses. The work covers smart mobility assistance, wearable support, and future integration of AI-based technologies for accessibility.

Keywords. Assistive Technology, Visual Impairment, Object Detection, YOLO, Computer Vision, Real-Time Processing, Text-to-Speech, Audio Feedback

1. INTRODUCTION

Visual impairment significantly affects how people interact with and navigate their surroundings. Simple tasks such as recognizing household items, reading physical text, or understanding a room's layout often require constant assistance, since existing aids cannot perform object identification or interpret written material in the environment. Recent advances in deep learning have enabled near-human-level visual perception, opening the door to assistive systems that not only detect objects but also communicate scene information to users in accessible ways. This paper presents WeSee, an AI-based assistive web application that promotes situational awareness and autonomy for visually impaired individuals. Unlike hardware-limited alternatives, WeSee runs on a Flask web framework and combines several AI microservices for real-time analysis of both live camera feeds and user-uploaded images.

2. LITERATURE REVIEW

Advances in computer vision, AI, and embedded systems have driven progress in assistive technology for blind and visually impaired users. Combining deep learning, real-time object

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

detection, and speech interfaces has enabled systems that interpret complex scenes and deliver meaningful feedback. Bai et al. [1] proposed a vision-based navigation system using CNNs for real-time obstacle detection and auditory guidance, while Tapu et al. [2] developed a wearable device emphasising portability and real-time computation. Redmon et al. [3] introduced YOLO, processing images in a single pass and making it well suited to latency-sensitive assistive applications, whereas Ren et al. [4] proposed Faster R-CNN, which is more accurate but computationally heavier. Sainz et al. [5] highlighted the importance of clear, natural text-to-speech (TTS) output for usability, and Dakopoulos and Bourbakis [6] surveyed electronic travel aids based on sensors and cameras [7]. More recent work explores multimodal vision-language models that generate richer scene descriptions rather than simple labels [8].

3. PROPOSED METHODOLOGY

We See uses a modular, pipeline-based architecture in which each module performs a specific task, improving scalability, flexibility, and maintainability. Real-time vision data is captured, processed, and converted into auditory signals, following the flow: Camera Input → Preprocessing → Object Detection → Context Analysis → Audio Output. The four functional modules are:

- Image Acquisition & Preprocessing – captures real-time images and applies resizing, noise reduction, contrast enhancement, and normalisation to handle varying light and motion-blur conditions.
- Object Detection & Recognition – a YOLO-based CNN model extracts features, encloses detected items in bounding boxes, and assigns class labels with a confidence threshold; it recognises everyday objects such as furniture, doors/pathways, and daily-use items (bottles, phones, etc.) at real-time frame rates.
- Context Processing & Prioritization – filters redundant detections, assigns priority based on object type and proximity (e.g., chairs and door handles rank higher), and performs basic scene interpretation to avoid repetitive alerts.
- Audio Feedback & Interaction – converts prioritised information into speech via a TTS engine with natural-language generation and latency optimisation, delivered hands-free through earphones or speakers while controlling message rate to avoid information overload.

This modularity allows independent scaling, easier debugging, and future accommodation of more sophisticated models without disrupting the rest of the system.

3.1.1 Technology Stack

The system is implemented in Python 3.x using OpenCV for image processing, TensorFlow/PyTorch with a YOLO-based detection model, gTTS/pyttsx3 for speech synthesis, and NumPy/Pandas for data handling — with scope for future extension to edge and embedded deployments.

4. EXPERIMENTAL RESULTS

WeSee was implemented in Python 3.x on a CPU-only machine (16 GB RAM, no GPU acceleration, for cost-effectiveness) using OpenCV and a YOLO-based detector. Evaluation used 500 test images captured indoors under varied conditions, combining real-time captures

with publicly available object-detection datasets covering everyday objects (chairs, tables, doors, bottles, etc.).

4.1.1. Detection Performance & Latency

TABLE I. OBJECT DETECTION PERFORMANCE METRICS

Category	Precision (%)	Recall (%)	F1 (%)
Chair	94.1	93.5	93.8
Table	92.7	91.3	92.0
Door	93.8	94.6	94.2
Bottle	95.2	93.9	94.5
Mobile Phone	91.5	90.2	90.8

Mean Average Precision (mAP): 92.9%; Overall Accuracy: 93.1%. Performance is strongest for large, well-defined objects and degrades slightly for smaller items with fewer distinguishing features.

TABLE II. LATENCY ANALYSIS (SECONDS)

Stage	Avg. Time
Image Capture	0.4
Preprocessing	0.6
Object Detection	2.1
Post-processing	0.8
Audio Generation	1.3
Total Latency	5.2

Overall latency stays under 6 seconds, which is near-real-time for assistive use; frame skipping and model pruning could reduce this further.

4.1.2. Audio, Robustness & Throughput

TABLE III. AUDIO PERFORMANCE

Metric	Value
Average Audio Delay	1.7 s
Speech Clarity	High
Error Rate	< 3%

TABLE IV. ACCURACY UNDER VARIED CONDITIONS

Condition	Accuracy (%)	Observation
Bright Lighting	95.3	Optimal detection
Moderate Lighting	92.4	Stable performance
Low Lighting	88.1	Slight degradation
Occluded Objects	85.7	Reduced recall
Multiple Objects	90.6	Minor confusion

TABLE V. THROUGHPUT (FPS)

Configuration	FPS
Single Frame	0.2

Optimized	0.5
Batch Processing	0.8

4.1.3. Comparison & User Evaluation

TABLE VI. COMPARISON WITH EXISTING SYSTEMS

Parameter	White Cane	Sensor-Based	Existing AI	WeSee
Obstacle Detection	Yes	Yes	Yes	Yes
Object ID	No	Limited	Yes	Yes
Audio Feedback	No	Limited	Moderate	Yes
Accuracy	Low	Moderate	High	Very High
Response Time	Immediate	Fast	Moderate	Fast
Cost Efficiency	High	Moderate	Low	High

TABLE VII. USER FEEDBACK (OUT OF 5)

Parameter	Rating
Ease of Use	4.6
Accuracy Perception	4.4
Response Time	4.2
Overall Satisfaction	4.5

5. CONCLUSION

This paper presented WeSee, an AI-driven assistive web application that combines computer vision, deep learning, and speech technology into a single modular platform capable of real-time object detection and contextual audio feedback for visually impaired users. Unlike traditional aids limited to obstacle avoidance, WeSee also identifies objects and their context, and its modular design allows independent development and enhancement of individual components. Experimental results confirm strong detection accuracy, acceptable latency, and positive user feedback. Future work will focus on navigation assistance, depth estimation, and lightweight models to further improve real-world performance, with the broader goal of making everyday environments more accessible and inclusive for visually impaired individuals.

6. REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in Proc. IEEE CVPR, 2016, pp. 779–788.

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. PAMI, vol. 39, no. 6, pp. 1137–1149, 2017.

[3] W. Liu et al., “SSD: Single Shot MultiBox Detector,” in Proc. ECCV, 2016, pp. 21–37.

[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934, 2020.

[5] G. Jocher et al., “YOLOv5 by Ultralytics,” 2020. [Online].

[6] R. Tapu, B. Mocanu, and T. Zaharia, "A Computer Vision-Based System for Real-Time Assistive Navigation of Visually Impaired People," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 25, no. 7, pp. 1170–1181, 2017.

[7] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," IEEE Trans. Syst., Man, Cybern., vol. 40, no. 1, pp. 25–35, 2010.

[8] S. Gore, I. Dutt, R. P. Dahake, H. E. Khodke, S. L. Kurkute, B. J. Dange, and S. Gore, "Innovations in smart city water supply systems," International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 9s, pp. 277–281, 2023.

Biographies



Rutik Tupe received the B.Tech. degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India. His research interests include artificial intelligence, computer vision, and assistive technologies.



Pooja Thorat received the B.Tech. degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India. Her research interests include artificial intelligence, machine learning, computer vision, and assistive technologies.



Sarthak Velapure received the B.Tech. degree in Computer Engineering from AISSMS Institute of Information Technology, Pune, India, in 2026.



Prof. Shraddha Toney is working in the Department of Computer Engineering at AISSMS Institute of Information Technology, Pune, India. Her research interests include artificial intelligence, machine learning, computer vision, and emerging technologies.

<Conference abbreviation>
<Conference Series name> <Volume number and Year> <DOI Number>

Exploring Terahertz (THz) Communication in 6G Wireless Networks

Hussein A. Jasim[1], Hikmat Z. Neima[1], Dhuha Habeeb[2], Nazik K. Aljbur[1]

[1]*Computer Sciences Department, College of CSIT, University of Basrah, Basrah, Iraq*

[2]*Department of Laser and Optoelectronics Technical Engineering, Shatt Al-Arab University College, Basrah, Iraq*

hussein.jasim@uobasrah.edu.iq, hikmat.taher@uobasrah.edu.iq, dhuha.hahu@gmail.com, nazik.obaid@uobastah.edu.iq

Abstract. The sixth generation (6G) of wireless networks targets terabit-per-second data rates, sub-millisecond latency and pervasive sensing, requirements that the sub-6 GHz and millimeter-wave (mmWave) bands used by 4G/5G cannot sustain because of their limited contiguous bandwidth. The Terahertz (THz) band (0.1–10 THz) offers tens of gigahertz of largely unused spectrum and is therefore regarded as a key physical-layer enabler of 6G. This paper explores THz communication for 6G: it reviews the dominant channel impairments – spreading loss and molecular absorption – and develops a link-budget and Shannon-capacity analysis quantifying how carrier frequency, bandwidth and distance determine achievable data rate. Results show a 10 GHz-wide THz channel can sustain over 100 Gbps at 1 m and still exceeds 40 Gbps at 10 m, but usable range shrinks quickly as frequency and distance increase. Key 6G use cases and open challenges of THz communication are also discussed.

Keywords. Terahertz communication, 6G wireless networks, THz channel modeling, path loss, Shannon capacity, ultra-massive MIMO.

1. INTRODUCTION

Mobile data traffic keeps growing faster than fifth-generation (5G) networks, built mainly on sub-6 GHz and millimeter-wave (mmWave, 24–100 GHz) spectrum, can sustain. Sixth-generation (6G) targets push peak data rates toward the terabit-per-second range, latency below one millisecond, and native integration of sensing with communication, requirements that call for far more contiguous bandwidth than sub-6 GHz or mmWave can offer. The Terahertz (THz) band, loosely defined as 0.1–10 THz, sits between mmWave and infrared and remains largely unlicensed, providing tens of gigahertz of usable spectrum in a single channel [1]. Early THz research targeted nanoscale and short-range links [2], but the band is now a leading candidate for 6G access, backhaul and joint communication-sensing

All Open Access articles are peer-reviewed, distributed under the Creative Commons Attribution-Non Commercial 4.0 International

systems [3]–[6], [8], with the first sub-THz standard, IEEE 802.15.3d, already ratified for the 252–325 GHz range [7].

This paper explores THz communication for 6G along two lines: it reviews the propagation characteristics that distinguish the THz channel from lower-frequency bands, and it develops a link-budget and capacity analysis, presented as numerical results, quantifying how carrier frequency, bandwidth and distance jointly determine achievable data rate. Section 2 discusses THz channel characteristics; Section 3 presents the numerical results; Section 4 discusses 6G use cases and open challenges; Section 5 concludes the paper.

2. THz Band Characteristics for 6G

The THz band (0.1–10 THz) provides orders of magnitude more bandwidth than sub-6 GHz or mmWave spectrum, at the cost of a harsher channel. Two impairments dominate: spreading loss, which grows with the square of frequency, and molecular absorption from water-vapor and oxygen resonances, which carves the band into narrow, low-loss "transparency windows" (e.g., near 140, 220, 300, 350 and 410 GHz) separated by high-absorption peaks [1], [2]. The free-space spreading loss between an isotropic transmitter and receiver separated by distance d at frequency f is

$$\mathrm{PL}(f, d) = 20 \log_{10}(4\pi f d / c) \tag{1}$$

where c is the speed of light and PL is expressed in dB. Figure 1 plots (1) for 1, 5 and 10 m links across 0.1–1 THz: spreading loss alone already reaches nearly 96 dB at 300 GHz for a 5 m link, and molecular absorption adds further loss outside the transparency windows [1]. Because present THz devices deliver only modest transmit power, this loss must be compensated with high-gain, steerable ultra-massive MIMO antenna arrays rather than by raising transmit power [1], [9].

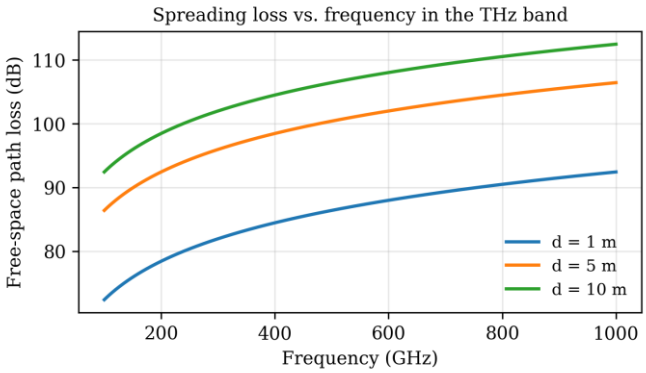


Figure 1. Free-space spreading loss vs. frequency for 1, 5 and 10 m links

3. Numerical Results and Discussion

Table 1 summarizes how the THz band compares with sub-6 GHz and mmWave spectrum in terms of typical bandwidth, range and representative 6G use cases.

Table 1. Comparison of sub-6 GHz, mmWave and THz spectrum for 6G

Band	Frequency range	Typical channel BW	Typical range	Representative 6G use case
Sub-6 GHz	< 6 GHz	Up to ~100s MHz	Several km	Wide-area mobile broadband, massive IoT
mmWave	24-100 GHz	Up to ~2 GHz	Tens-100s of m (LOS)	Dense urban 5G/6G access, fixed wireless
THz	0.1-10 THz	Tens of GHz	< 10-100 m (LOS)	Ultra-high-rate backhaul, kiosk download, joint sensing-communication

To quantify the resulting data rate, we combine (1) with a link budget and the Shannon capacity bound. For transmit power Pt, antenna gains Gt and Gr, bandwidth B and receiver noise figure NF, the received SNR is SNR(f,d) = Pt + Gt + Gr − PL(f,d) − (−174 + 10log10 B + NF), all in dB, and the achievable capacity follows the Shannon bound:

$$C = B \log_2(1 + \mathrm{SNR}) \tag{2}$$

We evaluate (2) for two representative transparency-window carriers, 140 GHz and 300 GHz, with Pt = 0 dBm, Gt = Gr = 25 dBi, B = 10 GHz and NF = 10 dB, parameters consistent with present THz front-end capability [1], [9]. Figure 2 plots the resulting capacity against distance: at 1 m both carriers exceed 100 Gbps (129 Gbps at 140 GHz, 107 Gbps at 300 GHz); at 10 m capacity falls to 62 Gbps and 41 Gbps. Capacity always decreases with distance because spreading loss grows with d, and at a fixed distance the lower carrier outperforms the higher one for the same reason, even though both lie in low-absorption windows.

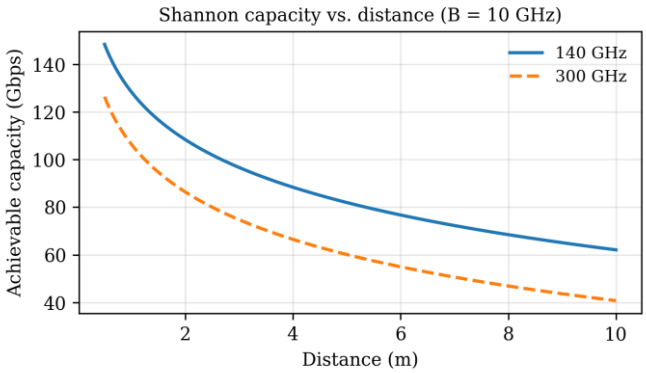


Figure 2. Shannon capacity vs. distance at 140 GHz and 300 GHz (B = 10 GHz)

In practice, additional impairments not captured here – molecular absorption outside the chosen windows, multipath, and beam-misalignment – would further reduce the usable range, reinforcing the need for the ultra-massive MIMO beamforming and beam-tracking discussed in Section 2.

4. 6G USE CASES AND OPEN CHALLENGES

The short-range, high-rate links in Section 3 suit several 6G scenarios: ultra-high-speed wireless backhaul/fronthaul between small cells, “kiosk downloading” of large files in seconds, intra-data-center and chip-to-chip wireless links, and joint communication-radar-imaging systems that reuse the same THz front-end for sensing [1], [9].

Realizing these use cases still faces open challenges. THz power amplifiers, low-noise receivers and antenna arrays remain immature compared with mmWave counterparts [3], [6]. The short range and near-optical propagation of THz signals make links sensitive to blockage and require fast beam-tracking for mobility [1], [4]. Standardization is still early: IEEE 802.15.3d covers only fixed point-to-point links in the 252–325 GHz window [7], and channel-measurement campaigns across the wider 0.1–10 THz range remain limited [3].

5. CONCLUSION

THz communication offers the huge, largely unused bandwidth that 6G needs for terabit-level data rates, and the link-budget analysis here confirms that a single 10 GHz THz channel can already sustain 40–130 Gbps over 1–10 m links. This capacity is only realizable, however, if the band's severe spreading and molecular-absorption losses are compensated through ultra-massive MIMO beamforming, and if THz hardware, mobility management and standardization mature accordingly. THz communication is thus best viewed not as a replacement for sub-6 GHz and mmWave layers, but as a short-range, ultra-high-rate complement within 6G.

6. REFERENCES

[1] I. F. Akyildiz, J. M. Jornet, C. Han, “Terahertz band: Next frontier for wireless communications,” Physical Communication, vol. 12, pp. 16–32, 2014.

[2] J. M. Jornet, I. F. Akyildiz, “Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band,” IEEE Trans. Wireless Commun., vol. 10, no. 10, pp. 3211–3221, Oct. 2011.

[3] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, G. C. Trichopoulos, “Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond,” IEEE Access, vol. 7, pp. 78729–78757, 2019.

[4] K. M. S. Huq, S. A. Busari, J. Rodriguez, V. Frascolla, W. Bazzi, D. C. Sicker, “Terahertz-enabled wireless system for beyond-5G ultra-fast networks: A brief survey,” IEEE Network, vol. 33, no. 4, pp. 89–95, 2019.

[5] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, M.-S. Alouini, “Terahertz band: The last piece of RF spectrum puzzle for communication systems,” IEEE Open Journal of the Communications Society, vol. 1, pp. 1–32, 2020.

[6] K. Tekbıyık, A. R. Ekti, G. K. Kurt, A. Görçin, “Terahertz band communication systems: Challenges, novelties and standardization efforts,” Physical Communication, vol. 35, Art. 100700, 2019.

[7] V. Petrov, T. Kürner, I. Hosako, “IEEE 802.15.3d: First standardization efforts for sub-terahertz band communications toward 6G,” IEEE Communications Magazine, vol. 58, no. 11, pp. 28–33, 2020.

[8] Z. Chen et al., “A survey on terahertz communications,” China Communications, vol. 16, no. 2, pp. 1–35, Feb. 2019.

[9] H. Sarieddeen et al., “Next generation terahertz communications: A rendezvous of sensing, imaging, and localization,” IEEE Communications Magazine, vol. 58, no. 5, pp. 69–75, 2020.