
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 System using an improved Random Forest framework. The proposed system incorporates hybrid feature selection to address dimensionality, learning drift to address changing patterns of attacks, and Explainable AI to enhance decision-making. The

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:

$$G = 1 - \sum_{i=1}^C P_i^2 \quad (2)$$

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 datasets, outperforming the existing works through improved 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

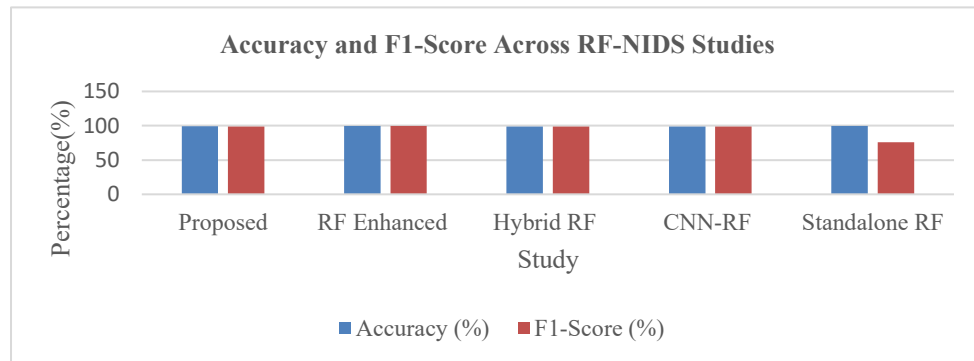


Figure 4.1. Accuracy and F1-Score Across RF-NIDS Studies.

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 nature of cyber threats, specifically the poor detection of rare classes of attacks and class imbalance problems 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 accuracy 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. Abdulhabeab, "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.