International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.052

Diagnosis of Heart Disease Using IoT in Edge with Multiple Machine Learning Techniques

Blessed Sam B
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur Chennai, India
bb4439@srmist.edu.in

Dr.Prakash M
Department of Data Science and Business Systems
SRM Institute of Science and Technology
Kattankulathur Chennai, India
prakashm2@srmist.edu.in

***Abstract*— Edge computing bridges the gap between industrial clouds and field devices for the Industrial Internet of Things (IoT). The most difficult aspect of edge computing is gathering data from different devices. The role of Edge is to process client's data at the network's edge, closer to the source. Due to the rapid increase in data traffic exchanged worldwide, there is an increasing need to collect and process data from sensors and devices of the IoT, which are operated in real time from remote locations and hostile operating environments. The proposed system aims to develop an edge-based heart disease prediction. The system was developed using a simple microcontroller that integrates temperature, heart rate and blood pressure. The measured value is tested with machine learning model to understand the patient's condition. Many feature combinations are used to reveal the prediction model and also for large number of well-known machine learning classification algorithms. Prediction model achieved the accuracy measure of 86% for SVM,92.24% for Decision Tree Classifier, 93.01% for Random Forest and 86.91% for K-Nearest Neighbor. So, from these four algorithms, Decision Tree Classifier gives the superior prediction, satisfying all the measures of precision, recall, accuracy and F-Score.**

Keywords—Edge Computing, IoT, Machine Learning, Performance measures, Heart Disease Prediction.

I. INTRODUCTION

Heart attacks and strokes were liable for 80% of people's death. The people who suffer from uncertain heart diseases, can be saved if it is predicated ahead of time. The (IoT) has already been widely used in a range of clinical aspects to capture the sensor values in attempt to diagnose the specific issues and predict cardiac problems through Machine Learning algorithms in real world applications and all healthcare standards. The IoT connects objects within the vicinity, collects the data from the people, and analyses it for early prediction and diagnosis[1]. Multiple inference has examined the capabilities of cloud-based and edge-based computing, concluding that the only system can manage the current scenarios in well-defined manner is Edge-based Machine Learning that can only meet the network delay, bandwidth, speed and all efficiency requirements. A superior execution of edge computing over traditional cloud computing can benefit many sectors, including healthcare[2].This paper's contributes, cardiac patient monitoring system based on the concept of IoT, using various signal sensors and a microcontroller. IoT technology is increasingly being used by sensor networks to capture, interpret, and transmit data from one node to another. IoT is a relatively new and rapidly expanding technology in which multiple sensors can sense, share, and communicate over a local network, public networks and through Internet Protocol. After a specified amount of time, the sensors collect data, analyse it, and use it to initiate the necessary action, resulting in an cloud-based network intelligence foranalysis, planning, and decision making. IoT-enabled products, such as embedded technology, enable data exchange. So here the methodology was tested on only three subjects and predicted their condition. So here the key contribution is to select a machine learning model that which best suits for the prognosis of heart disease.

II. RELATED WORKS

In this literature, various number of sensor-based wearable health monitoring devices in machine learning with edge has been suggested. This section discusses a few related works in detail. Every year, chronic diseases such as cardiovascular disease, respiratory disorders, and brain disorders kill more than 65% of the world's population. Chronic disease patients have a long recovery period and require technology for consistent monitoring of their health condition. Wearable technology is used to monitor vital parameters such as blood pressure, Oxygen saturation, respiratory rate, and Electrocardiograph for the patients.

Ram et al[2022] Machine learning in edge has been investigated for detecting anomalies and improving the accuracy of the data and mobile health monitoring can predict outcomes. Moreover, multi-modal sensor information was pre-processed to cleanse the data and classify the occurrences in the dataset[3].

Md.MahbuburRahman et al[2022] Data that has been pre- processed in order to train and evaluate machine learning algorithms. The first stage divides pre-processed data into two portions. The majority of these are utilised in the training phase (80%), with the remainder (20%) employed in the testing phase. dataset was trained using machine learning techniques such as Decision Tree, XgBoost, KNN, Support Vector Machine[4]

Ayushi Das et al[2022] Four different classification algorithms to forecast certain widely identified diseases are k-Nearest Neighbour, Naive Bayes, Decision Tree, and Random Forest. On a disease prediction data-set, these Supervised Machine Learning classifiers are used to identify 41 prevalent diseases based on any 5 prominent symptoms from the dataset's 132 common symptoms[5].

JameelAhamed et al[2021] To examine the available data on cardiovascular disorders in order to predict and prevent heart disease at an earlier stage. The heart disease patient dataset was collected and stored in the cloud. The information stored is then pre-processed and examined

DOI: rp-9788770040723.052

further using machine learning techniques to forecast cardiac problems[6].

ButchiRaju et al[2021] This model combines Edge-Fog- Cloud computing to deliver accurate and timely results. The hardware components collect information from various patients. To obtain significant features, heart feature extraction from signals is performed. In addition, with, thefeature extraction of other attributes are also combined. All of these characteristics are gathered and fed into the diagnostic system via an Optimized Cascaded Convolution Neural Network[7].

N.Saranya et al[2020] Latency reduction approaches in mobile computing and fog networks, data availability, power computation, and other parameters are analysed and compared to see which gives superior results[8].

R. Latha et al[2019] The blood viscosity is predicted by Partially Observable Markov Decision Process. The decision was made to address two causes of ambiguity. The first is about control, specifically diagnosis. The other is related to the user, namely the decision maker. A hierarchical dynamic effective framework model was chosen to solve this POMDP. Structure-based approximations and approaches are used to reduce complexity[9].

S. Mohan et al [2019] An innovative way for identifying key features by using machine learning techniques, with the goal of increasing accuracy of cardiovascular disease prediction. The prediction model is presented with several combinations of features and various recognised categorization algorithms. We achieve an improved performance level with an accuracy level of 88.7% using the hybrid random forest with a linear model to predict heart disease[10].

ShadmanNasif et al[2018] Cardiac patient monitoring system based on the Internet of Things, with various physiological sensors and an Arduino based microcontroller. Sensor networks are now using IoT technology to collect, analyse, and transmit data from one node to another node[11].

TABLE I.ONGOING HEALTH MONITORING DEVICE

S. No	Motive	Signals	Analysis (edge/cloud)
1	Observing heart illness	ECG, Heart rate, EMG, Accelerometer	Edge with ML
2	Heart Rate Monitor	ECG, EMG, BPV	Edge with ML
3	Tracking sleep	Temperature, Heart Rate	Cloud with ML
4	Cardio vascular disorders	ECG, Blood Pressure, SPO2	Cloud with ML
5	Heart disease Monitoring	ECG, EEG, EMG, Temperature, Oxygen level, Respiratory rate	Edge with ML

III. FINDINGS FROM RELATED WORKS

The majority of the research works analysed cardiovascular disease models using single and two algorithm combinations that are not powerful enough to provide reliable predictions in underdeveloped countries.

The IoT-ML-Edge paradigm is a distinct Information Technology (IT) prototype in which IoT, ML, and edge are three interconnected technologies integrated together to address both existing and future world difficulties associated to the health-care concern.

For cardiovascular disease prediction, we employed three distinct feature choices as well as four classifiers. We have yet to come across any research that has used such classifiers for feature selection and cardiovascular disease prediction. In most occurrences, research scientists employed just three algorithms in their work.

In this system we used various machine learning techniques to forecast cardiac illness because it is one of the best andmost novel tools for prediction work, and it is also used in other subjects such as cancer symptoms and weather prediction system.

IOT- Edge Module

Sensors

The sensor framework incorporates ECG to monitor the blood pressure and sugar, pulse sensor to measure the heartbeat and temperature sensor to monitor the patient's vital signs.

Edge module

The data's which are obtained from the sensor component will communicate to edge module. Sensor readings are captured and then updated in comma-separated value sheet. Then the samples from the files are sent to further processing for testing the data.

Processing Module

In this stage, Machine Learning algorithms are used to detect the abnormality of the patient's condition. So, this module is equipped with various types of algorithms to check whether the patients have disease or not.

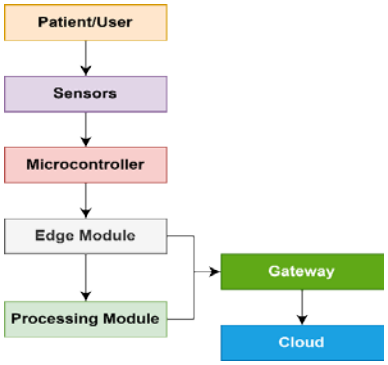


Fig. 1 Systematic representation of IoT Edge Module

III. PROPOSED SYSTEM

Edge technology push data processing closer to the edge of the network, leading to faster response time and greater efficiency. Instead of continually transferring data to the cloud for computing activities, which imposes energy expenses, data may be collected and processed locally to the user on edge devices and server. The Internet plays an essential role in the development of an IoT-based healthcare monitoring system. Keeping records for any doctor is tough due to the vast amount of data.

DOI: rp-9788770040723.052

Physicians, on the other hand, must utilize this prior data to forecast a patient's health state. Over the years, several machine learning technologies have been applied in the medical application field. This research work presents an edge-based cardiovascular related disease prediction using IoT. This wearable gadget has a temperature sensor, a heart rate sensor, and blood pressure sensor. The raspberry pi controller is linked to these sensors. Using a heart disease prediction data, the controller is trained with a machine learning model. Once if the data is obtained from individuals, the controller analyses and shows whether the patients have cardiac disease or not. So here we employedfour Machine Learning algorithm techniques to predicted the best outcome.

Architecture

1. The first stage is pre-processing of data, the patient data is collected through sensors, and the sensor input data is pre- processed. Data that is missed is deleted during pre- processing.
2. The second step involves data feature selection and heart disease features such as respiratory symptoms, cholesterol, and blood sugar level and these parameters are selected for further processing.
3. The third stage is data splitting. At this point, the patient's full data is divided into 70% and 30% for training and testing purposes.
4. So, in the final stage, the data is trained with various machine learning techniques then at last the testing of model is processed.

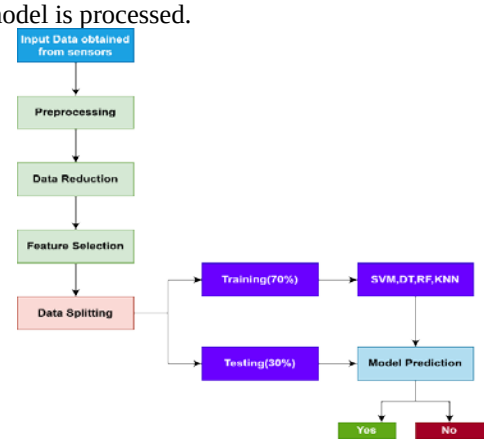


Fig. 2 Flow Diagram for Heart Disease Prediction.

V. Data Processing

Data Collection

The sensors capture the patients' data, as well as vital indicators such as ECG, pulse rate, and cholesterol.

Data Cleaning

The pre-processing approach known as data cleaning method is utilized here to clean the redundant data, noisy data and outliers.

Data Reduction

The feature selection mechanism is employed here to choose exactly the particular data needed for training the model.

Data Transformation:

Data transformation converts data from one pattern, standardized measures, or structure to another without affecting the data's content. So, in this case, the technique called data normalization strategy is used to eliminate the unstructured data.

Data Integration:

The method of incorporating data from numerous source systems in order to generate homogenous sets of information for operational purposes. Here the technique called ETL isused to Extract, Transform and Load the data from sensors and the Python script is used here to serially capture real- time data and log it into a CSV (Comma Separated Value) file.

VI. Methods

Support Vector Machine

SVM modelling is a good classification method for predicting individuals with heart failure. This prediction model aids in clinical diagnosis, allowing data decisions to be made and patients to be handled effectively. To generate more accurate prediction of heart disease this system adapts the SVM machine learning technique. The patient data includes the following attributes - age, gender, cholesterol, blood pressure and pulse rate. Fig.3 shows the output of support Vector Classifier scores for different kernels were linear kernel achieves a score of about 86%.

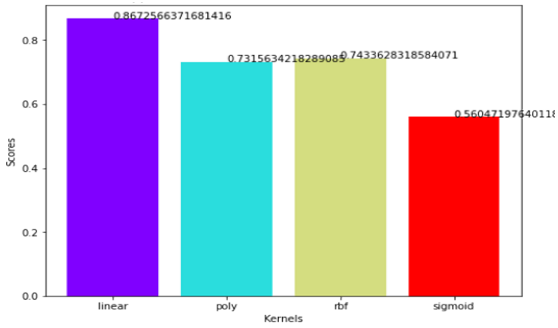


Fig. 3 Execution Graph for different Kernels in Python

TABLE II. Scaling Parameters

No	Kernel	Scores
1	Linear	0.86
2	Polynomial	0.73
3	RadialBasis Function	0.74
4	Sigmoid	0.56

Decision Trees

A decision tree is a machine learning method that may be applied for both regression and classification technique. It is a classifier with a hierarchical tree structure. It reveals useful hidden information. The massive data collection may also be used to generate new target patterns. For categorization, decision trees are functional in multiple disciplines such as machine learning and to extract the information's. A decision tree does not need additional subject expertise. It is simple to understand and fast. Decision trees may handle a variety of data kinds, including nominal, ordinal, binary, and real values[12]. It builds decision nodes in which the internal

DOI: rp-9788770040723.052

nodes represent data characteristics, the edges represent decision rules, and each leaf node represents the class label, i.e., the target class label is specified, if that defined path is taken. The step begins with the root node of the decision tree; subsequent nodes are picked, and the process continues until the lead node is reached; the class label at the leaf node is the predicted class label[13]. Fig.4 shows the decision tree classifier, the class obtains the value either HD (Heart Disease) or NHD (No Heart Disease) based on the conditions.

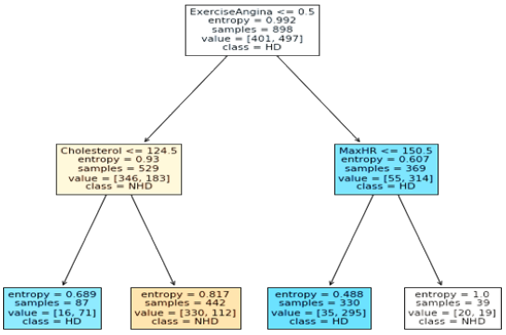


Fig 4. Execution of Decision Tree Classifier

Random Forest

It is a supervised machine learning approach that builds a series of decision trees. Majority of the decision trees are used to make the ultimate conclusion. Random Forest is used to create a classifier model capable of predicting illness with greater performance and accuracy. It is possible to cure them with appropriate therapies using modern medical technologies. However, if illness is detected late, even the most advanced medical technology cannot help. Fig.5 shows the actual values and the predicted values from test dataset.

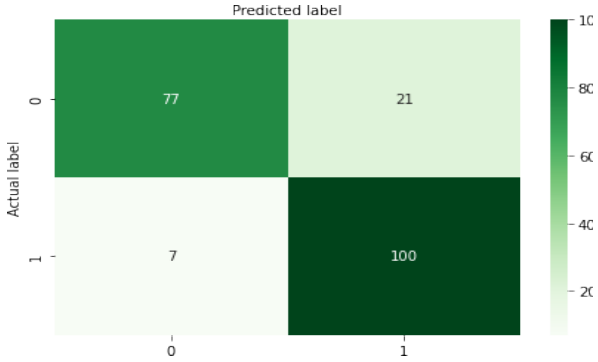


Fig 5. Confusion Matrix

K-Nearest Neighbor

KNN looks for similarities between predictions and data in the dataset. KNN employs a non-parametric system, since there is no significant discovery of parameters associated with a particular functional form. It does not make any speculations about the dataset's properties or output. As a result, majority of the computing effort happens during classification rather than training. KNN typically works attempting to determine the closest class feature, and then assigning it to that particular class which is nearest to that point, from fig.6, it's clear that minimum error rate is 0.105058 at K=3.

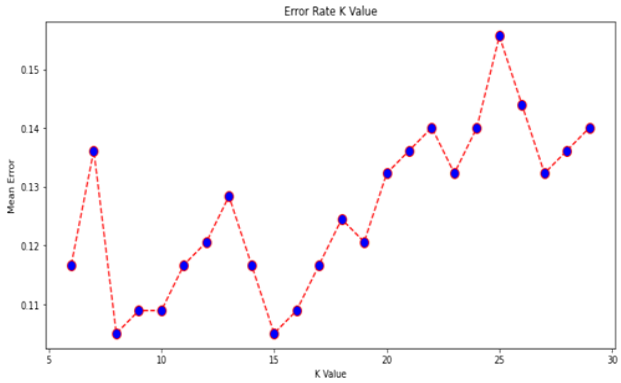
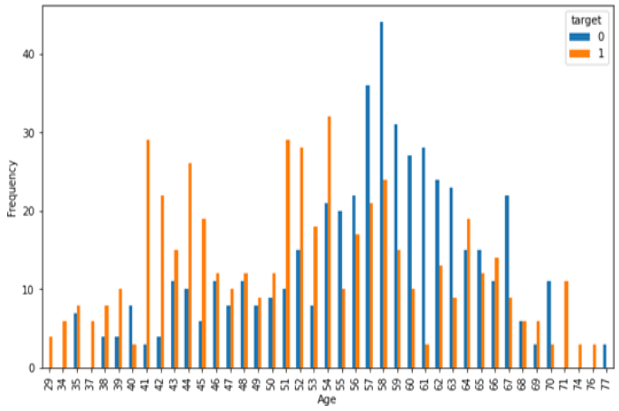


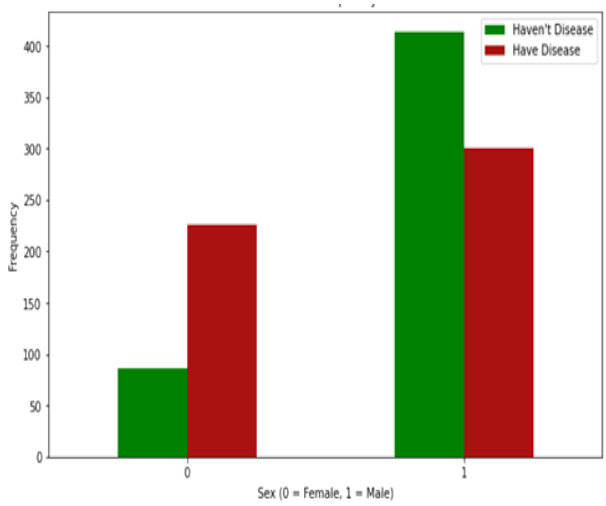
Fig 6. Minimum Error Rate

VII. Findings and Discussions

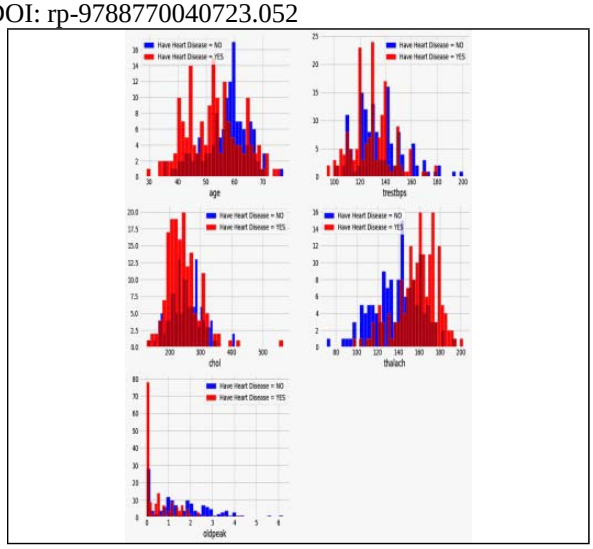
The goal of this research is to determine the fact that the patient has cardiac disease or not. The entries in the data are separated into two groups. (i.e., 70% for training and 30% for testing). This section outlines the categorization model results obtained by Python language. Fig shows the Heart Disease Prevalence by Age and Fig shows the heart disease frequency for male/female. Fig shows the probability of heart disease in terms of blood pressure, cholesterol and pulse rate. Fig shows the comparative analysis of various supervised learning algorithms. The best accuracy score is provided by random forest with 93.21%.



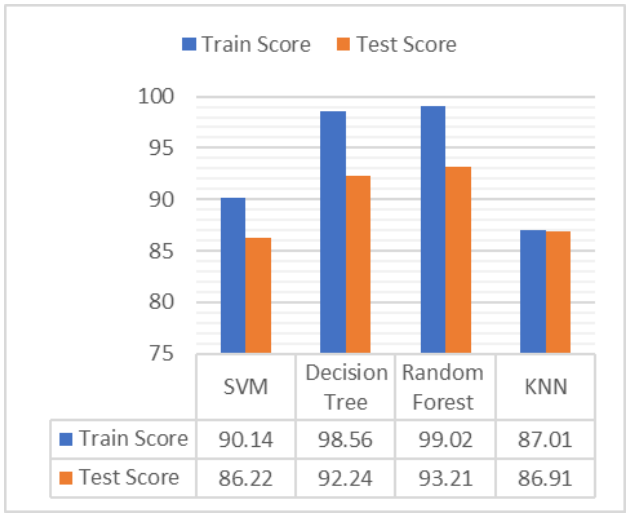
Output 1: Heart Disease Frequency (Ages)



Output 2: Heart Disease Frequency (Gender)



Output 3: Prediction Graph



AC Accuracy Chart

VII. Conclusion

Edge computing-based Applications have received considerable attention. The majority of present wearable health monitoring devices are designed to process data on the cloud. However, the purpose of this work is to experiment with an edge-based system for health monitoring. The purpose is to investigate and combine majority of data in order to generate a more meaningful dataset that includes a diverse variety of population patterns. For the prediction of heart disease, feature selection may be applied to extract more relevant characteristics and effective findings. This study compared four machine learning algorithms for heart disease prediction and found promising results. From this work, the performance metrics of Machine Learning algorithms are evaluated and found the best technique for the prognosis of heart disease.

VIII. Future work

Furthermore, data pre-processing approaches and machine learning classification algorithms can be used in future employment to get good outcomes than those achieved in this current work. In most cases the patient does not have sufficient time to go to the doctor, so in such cases, creatinga simple application for the user - interface through smart mobile to resolve this issue, and this service makes the prediction task quicker from the patient's location.

References

[1] A. Jenifer et al., “Edge-based Heart Disease Prediction Device using Internet of Things,” Proc. - Int. Conf. Appl. Artif. Intell. Comput. ICAAIC Icaaic, pp. 1500–1504, 2022, doi: 10.1109/ICAAIC53929.2022.9793104.

[2] P. Gupta, A. V. Chouhan, M. A. Wajeed, S. Tiwari, A. S. Bist, and S. C. Puri, “Prediction of health monitoring with deep learning using edge computing,” Meas. Sensors, vol. 25,November 2022, 2023, doi: 10.1016/j.measen.2022.100604.

[3] S. S. Ram, B. Apduhan, and N. Shiratori, A Machine Learning Framework for Edge Computing to Improve Prediction Accuracy in Mobile Health Monitoring, LNCS. Springer International Publishing, vol. 11621, 2019. doi: 10.1007/978-3-030-24302-9_30.

[4] M. A. Rahman and M. ShamimHossain, “An Internet-of-Medical-Things-Enabled Edge Computing Framework for Tackling COVID-19,” IEEE Internet Things J., vol. 8, no. 21, pp. 15847–15854, Nov. 2021, doi: 10.1109/JIOT.2021.3051080.

[5] A. Das, “A Collaborative Empirical Analysis on Machine Learning Based Disease Prediction in Health Care System in Health Care System,” 2022.

[6] J. Ahamed, A. M. Koli, K. Ahmad, M. A. Jamal, and B. B. Gupta, “CDPS-IoT: Cardiovascular Disease Prediction System Based on IoT Using Machine Learning,” Int. J. Interact. Multimed. Artif. Intell., vol. 7, no. 4, pp. 78–86, 2022, doi: 10.9781/ijimai.2021.09.002.

[7] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., &Verma, A. (2022). Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks. Journal of Sensor and Actuator Networks, 11(4), 82.

[8] “Literature Survey of Data Latency Reduction Techniques in Mobile Edge Computing-Iot,” J. Crit. Rev., vol. 7, no. 06, Apr. 2020, doi: 10.31838/jcr.07.06.151.

[9] R. Latha and P. Vetrivelan, “Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing,” 2019 11th Int. Conf. Commun. Syst. Networks, COMSNETS 2019, vol. 2061, pp. 833–837, 2019, doi: 10.1109/COMSNETS.2019.8711358.

[10] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications. Computers and Electrical Engineering, 102, 108130.

[11] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, “Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System,” World J. Eng. Technol., vol. 06, no. 04, pp. 854–873, 2018, doi: 10.4236/wjet.2018.64057.

[12] A. Kishor and W. Jeberson, “Diagnosis of Heart Disease Using Internet of Things and Machine Learning Algorithms,” Lect. Notes Networks Syst., vol. 203 LNNS, pp. 691–702, 2021, doi: 10.1007/978-981-16-0733-2_49.

[13] R. VijayaSaraswathi, K. Gajavelly, A. KousarNikath, R. Vasavi, and R. Reddy Anumasula, “Heart Disease Prediction Using Decision Tree and SVM,” no. February, pp. 69–78, 2022, doi: 10.1007/978-981-16-7389-4_7.