Internet of Things based Framework for Smart Healthcare Using Hybrid Machine Learning

Henry Okemiri (henry.okemiri@funai.edu.ng)
Alex Ekwueme Federal University Ebonyi state
Alo Uzoma
Alex Ekwueme Federal University Ebonyi state
Nnamene Christopher (nnamenechris@yahoo.com)
Alex Ekwueme Federal University Ebonyi state

Research Article

Keywords: Cardiovascular diseases, Internet of Things, Naïve Bayes, Random Forest, ROC Curve, Real-time Monitoring and Diagnosis, Support Vector Machine

Posted Date: March 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1428239/v1

Internet of Things based Framework for Smart Healthcare Using Hybrid Machine Learning

Henry A. Okemiri[1*], Rita U. Alo[1], and Christopher C. Nnamene[1]
[1] Department of Computer Science and Informatics, Alex Ekwueme Federal University, Ndufu-Alike, Nigeria
[e-mail: henry.okemiri@funai.edu.ng]
[*]Corresponding author: Henry A. Okemiri

Abstract

The world has been facing the challenge of increased cardiovascular death rate for decades now. The utilization of machine learning techniques coupled with the power of Internet of Things can be an effective solution to this real world problem. This research presents Internet of Things based framework for smart healthcare using hybrid machine learning for monitoring heart disease, a system specifically designed for wearable devices that facilitates heart diseases monitoring. The system uses wearable sensors to collect observable vital signs which is contextualized with data from clinical database for a high efficiency in analysis and prediction. Internet of Things and Machine Learning have gained wide applicability in the healthcare sector. The problem of inaccurate Electrocardiogram interpretation and high dependency on individual interpretation, coupled with delayed analysis and diagnosis of patients and lack of accurate prognosis are what this research solves. Machine learning was used accurate analysis and predictions of diseases while the internet of things components was communicating the server for smartness. The framework consists of some components like data collection using wearable sensors, data analysis center, which uses hybridized machine learning algorithms, to quickly identify potential cardiovascular cases from real-time symptom data. This facilitates real-time communication between doctor and patient through wearable technologies. With this, real-time analysis and monitoring of cardiovascular disease would be possible. This system instead of transferring the raw data collected from patients directly to the health care professionals, sends those data to a python platform for analysis on the local device by feeding the hybrid of collected data to Random Forest, Naïve Bayes and Support Vector Machine (SVM) algorithms to analyze and monitor features extracted from clinical databases and wearable sensors and then classify a patient as “Negative” or “Positive” for heart disease. In this research, Object Oriented Methodology and Analysis was used with arduino and python programming languages. The results of our experiment showed that the system was successful in both analysis and classification of a patient’s heart disease with an accuracy of 87.6% for the hybrid machine learning. Random Forest model has a precision and recall that exceeded 100%, 75% precision and 65% recall for Naïve Bayes model and 77% precision and 56%

recall for support vector machine. We used UCI clinical dataset of 1025 samples of patients with 25% for test data and 75% for the data training.

Keywords: Cardiovascular diseases, Internet of Things, Naïve Bayes, Random Forest, ROC Curve, Real-time Monitoring and Diagnosis, Support Vector Machine.

Introduction

In recent years, healthcare services have rapidly evolved to provide wireless communication media between doctor and patient via wearable devices, which is referred to as telemedicine. This allows for real-time monitoring of chronic illnesses such as heart failure, asthma, hypotension, and hypertension in areas where medical facilities are few, such as rural areas, or for people who have been off of health services for a while. Heart disease becomes the primary cause of death in all of these situations as a result of a shift in lifestyle that affects people of all ages. According to the literature, nearly 2.8 billion individuals die each year as a result of heart disease caused by being overweight or obese, which affects cholesterol levels, blood pressure fluctuations, and, most critically, the impact of stress hormones on long-term heart health. Many wearable technologies assessed common heart parameters such as blood pressure, blood glucose level, blood oxygen saturation, ECG, and so on. In light of all of this, the importance of hormonal imbalance as a result of stress, i.e. the person's mood (mental health status), and the impact of good vs. bad cholesterol are also discussed in depth [1].

In this era of communication and connectivity, individuals have multiple technologies to support their day-to-day requirements [2]. Due to this, Internet of Things (IoT) and Machine Learning (ML) are emerging technologies for practical solutions to problems facing several sectors, especially the healthcare sector. Machine learning and IoT work towards creating a better technology and environment, which will ensure efficiency and productivity for the healthcare sector [3]. IoT is a framework that uses technologies like sensors, network communication, artificial intelligence and bigdata to provide real life solutions. These solutions and systems are designed for optimal control and performance. In IoT, a single device can generate immense amounts of data every second. All these data from IoT devices are transmitted to servers or gateways to create better machine learning models [1]. This increased use of IoT yields huge amounts of raw data that are effectively processed by using machine learning to derive many useful insights that can become positive changers especially in healthcare delivery [4].

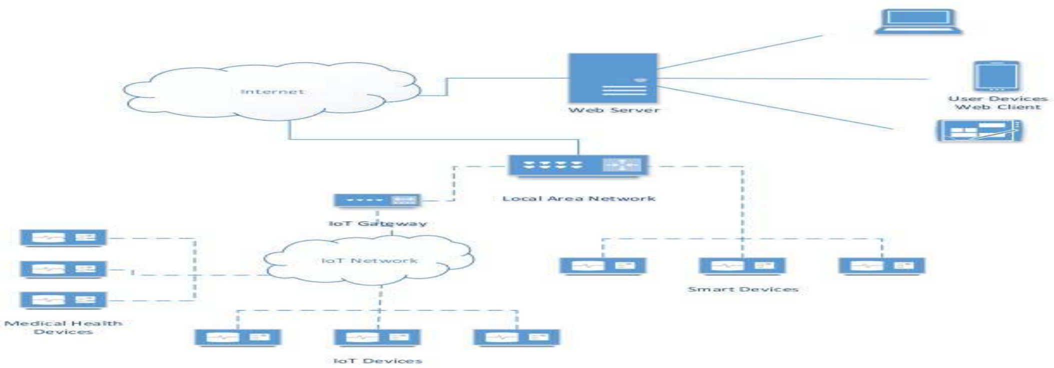


Fig. 1 - IoT Deployment and Application Scenarios [4]

A typical IoT network in a home is depicted in Fig. 1. Different IoT devices are shown in the diagram, either directly connected to an Internet router or via an IoT gateway, which functions as a bridge between the limited IoT network and the internet. IoT devices connect to the internet in this diagram to post data and receive requests from a server. Users can access and control data uploaded from their devices using a web interface and web apps installed on the server.

Internet of Things (IoT) as a global network of smart devices that can sense and interact with their environment for communication with users and other things (smart devices) and systems has played critical roles in enhancing healthcare delivery [5]. With increased attention on efficiency and better outcomes, the diffusion of IoT and ML could play an immensely important role in reducing healthcare costs without reducing the quality of care delivered to patients [6]. Although there are innovative implications of the integration of IoT in modern health applications, remote patient monitoring system that integrates IoT with ML has received less attention. However, IoT and ML have the potential to bring about new medical applications such as remote health monitoring, chronic disease, fitness programs, and adult care. Hence, medical devices including diagnostic and imaging devices and sensors can refer to smart devices representing the core of IoT [6].

IoT smart devices have limited resources, so it is necessary to consider other alternatives to analyze data, such as machine learning [7]. Machine learning is a subset of artificial intelligence that consists of studying the algorithms and statistical models used in computer systems in order to achieve specific objectives effectively, based on patterns and inferences [7]. In this context, there are several challenges in the health sector that provide areas of opportunity for the IoT and machine learning to provide solutions or alternatives that contribute to improving the healthcare or quality of life of patients. This work presents machine learning and IoT based smart health platform for the prevention, detection, treatment, and control of cardiovascular diseases, and other diseases or health problems derived from this condition. Machine learning techniques are essential in this work due to the need to classify patients based on their biomedical variables and behavior in order to provide the most appropriate recommendations to improving their health.

The purpose of this study is to underscore automated IoT based framework for remote patient monitoring using hybrid machine learning which will automatically trigger medical alerts, hospital visits, and medications in a timely and precise manner. Within this framework, patient data collected from various remote sensors and user devices are easily integrated, analyzed, and transformed into meaningful information for related users (e.g., physicians, nurses) to rapidly respond and take actions in accordance with their standard operating guidelines.

Statement of the Problem

The tracking and monitoring of patients and their healthcare actors remain a critical challenge in the healthcare sector. Medical errors remain third highest killer claiming over 400,000 lives per annum [8].

These errors include:

i. Delay in analysis and diagnosis of patients.

ii. Increase death rate due to heart failures in remote places
iii. Lack of accurate ECG interpretation and high dependence on individual interpretation.
iv. Inability of medical personnel's to keep track of changes in the health parameters of the patient over a period of time
v. Lack of accurate prognosis (prediction).

Objectives of the Study

The aim of this thesis is to develop an Internet of Things framework for smart healthcare using hybrid machine learning algorithm with the following objectives:

i. To develop a model for accurate detection and diagnosis of heart diseases.
ii. To hybridize clinical dataset and data collected by sensors from patients
iii. To implement a remote real-time patient analysis, diagnosis and monitoring
iv. To develop multiple parameters monitoring system for effective diagnosis.

Overview of Machine Learning and IoT Healthcare

The use of IoT-based devices is changing people's lifestyles, particularly in activities related to healthcare. In this sense, IoT-based devices monitor, analyze, diagnose, and contribute to the generation of medical recommendations for various health conditions, such as heart disease, overweight, obesity etc. For this reason, this topic has become the focus of much attention in recent research. In this section, we present a review of the state of the art of research involving the ML and IoT in healthcare, particularly with respect to cardiovascular diseases.

Neural networks are generally regarded as the best tool for prediction of diseases like heart disease and brain disease. A method which uses 13 attributes for heart disease prediction. The results showed an enhanced level of performance compared to the existing methods in works like (Al-milli, 2013). The Carotid Artery Stenting (CAS) has also become a prevalent treatment mode in the medical field during these recent years. The CAS prompts the occurrence of major adverse cardiovascular events (MACE) of heart disease patients that are elderly. Their evaluation becomes very important. Results were generated using an Artificial Neural Network ANN, which produces good performance in the prediction of heart disease [10], [11]. Neural network methods combine not only posterior probabilities but also predicted values from multiple predecessor techniques. This model achieves an accuracy level of up to 89:01% which is a strong result compared to previous works.

Machine learning approaches have already been used to characterize and predict a variety of health risks [12]. Recent works identify patients with undiagnosed peripheral artery disease and predict their mortality risk found that such an approach outperforms a simpler stepwise logistic regression in terms of accuracy, calibration, and net reclassification [13]. Such predictive models have been implemented in medical practice, resulting in more efficient and better quality care. In a related work, [14] described a framework designed specifically for the healthcare industrial IoT (HealthIIoT), in which information was obtained through sensors and smart devices. Likewise, [15] presented research to create procedures and instruments that benefit semantic interoperability in mobile health through the "INTER-IOT" project. [15] provided a general overview of tele-health and considered new tele health technologies and tools to increase the quality of healthcare services. In addition, [16] described future trends in wireless communication with a focus on 5G networks, in which the benefits for the IoT and e-health are notable.

In [10], Artificial Neural Network (ANN) produceed the highest accuracy prediction in the medical field. The back propagation multilayer perception (MLP) of ANN is used to predict heart disease. The obtained results are compared with the results of existing models within the same domain and found to be improved [17]. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with neural network, decision tree, Support Vector machines (SVM), and Naive Bayes. The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 86:8% for *F*-measure, competing with the other existing methods. The classification without segmentation of Convolutional Neural Networks (CNN) is introduced. This method considers the heart cycles with various start positions from the Electrocardiogram (ECG) signals in the training phase. CNN is able to generate features with various positions in the testing phase of the patient [18], [19]. A large amount of data generated by the medical industry has not been used effectively previously. The new approaches presented here decrease the cost and improve the prediction of heart disease in an easy and effective way. The various different research techniques considered in this work for prediction and classification of heart disease using ML and deep learning (DL) techniques are highly accurate in establishing the efficacy of these methods [20], [21].

In a similar work, [22] intends that large data existing from medical diagnosis is scrutinized by means of data mining tools and valuable information known as knowledge is hauling out. Mining is a method of investigating colossal sets of data to acquire the patterns which are hidden and formerly unknown associations and knowledge detection to facilitate the enhanced understanding of medical data to thwart heart disease. There are several DM techniques available namely Classification techniques concerning Naïve Bayes (NB), Decision tree (DT), Neural network (NN), Genetic algorithm (GA), Artificial intelligence (AI) and Clustering algorithms like KNN, and Support vector machine (SVM) [16]. In recent years, many machine learning algorithms have been applied in ECG classification, such as Neural Network, K-Nearest Neighbors algorithm, and Support Vector Machine (SVM). Although these studies provide solutions in detecting cardiac diseases, they did not propose the hardware of the ECG acquisition and transmission for wireless applications [23].

The expectation of cardiovascular illness by methods for a few machine learning calculations is going on [24]. Many research papers have executed different machine learning calculations, for example, Naive Bayes, Random Forest, Gradient boosting, Logical Regression and Support Vector Machine for anticipating cardiovascular illness [1]. In [25] it depicts how these machine learning calculations are utilized to foresee the pneumonia ailment.

We have also seen recent developments in machine learning ML techniques used for Internet of Things (IoT) as well [26]. ML algorithms on network traffic data has been shown to provide accurate identification of IoT devices connected to a network. [26] collected and labeled network traffic data from nine distinct IoT devices, PCs and smartphones. Using supervised learning, they trained a multi-stage meta classifier. In the first stage [31], [32] and [33], the classifier can distinguish between traffic generated by IoT and non-IoT devices. In the second stage, each IoT device is associated with a specific IoT device class. Deep learning is a promising approach for extracting accurate information from raw sensor data from IoT devices deployed in complex environments [27], [28]. Because of its multilayer structure, deep learning is also appropriate for the edge computing environment [29] and [30]

Heart rate monitoring using pulsed PPG signals during intensive exercise is a subject studied in [34] and [35]. As the signals get corrupted due to extreme movement disorders caused by abrupt hand movements, [32] proposed and compared with techniques used, a general framework called TROIKA, comprising a signal decomposition for the dismemberment, signal reconstruction spectrum, spectral analysis, and verified spectral data analysis. Already in [34], a new technique was presented to accurately determine the heart rate during excessive movement by classifying PPG signals obtained from smart- phones or wearable devices combined with motion data obtained from accelerometer sensors. The approach uses the IoT cloud connection from smart phones to PPG signal selection using deep learning.

However, the presence of noise primarily due to motion strongly affects the result of the PPG analysis. The study in [36] states in its proposed method that physiological signal involves preprocessing, specifically the breakup, and can substantially improve the overall performance efficacy and clinical utility as demonstrated in the case study, which shows a significant improvement in efficiency when identifying coronary artery disease (CAD) from the PPG signal. The increased popularity of visible technologies opened the door to IoT-based solutions for health care. Portable health-tracking devices are excellent candidates to minimize the distance between the patient and the doctor. One of today's most prevalent health problems is the low survival rate of sudden cardiac arrests. All existing systems for predicting cardiac arrest primarily consider heart rate parameters. Some research ad- dresses this problem by concentrating on the analysis and detection of ECG signals that eventually lead to a risk forecast, as abnormal ECG patterns indicate potential heart [37].

IoT-Based Machine Learning Approaches in Healthcare

A smart personal health advisor (SPHA) is proposed in [38], which is used to assess the general health conditions of users. Physiological conditions may be analyzed from body parameters, such as blood pressure, body temperature, and heartbeat, which are mostly obtained from (i) structured data—user's medical record [39] (ii) text data—doctors' diagnoses [40] and (iii) video and image data—obtained from medical devices [41]. The proper recognition of psychological conditions is difficult and challenging, therefore, it is obtained from psychological signs, such as voice and facial expressions, through temporal Bayesian fusion [42]. In this technique, the data is combined and then on the basis of that data, psychological characteristics of users can be extracted, and hence the condition is recognized. This idea is depicted in Fig. 3.

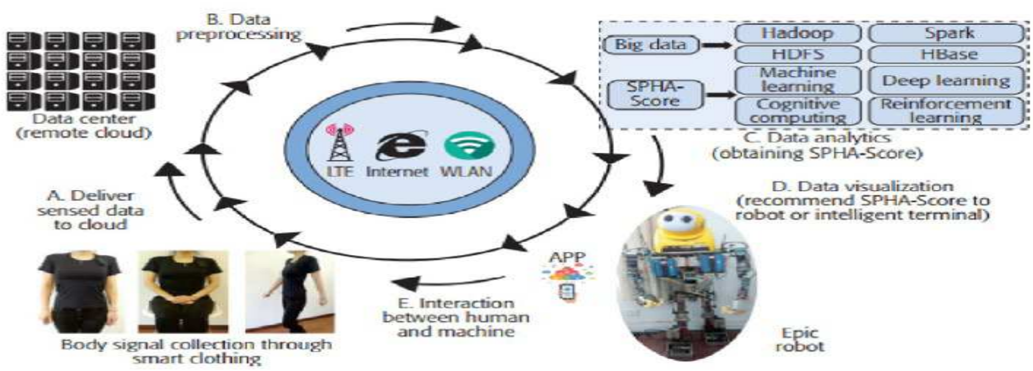


Fig. 3. SPHA scenario. Source: [43].

The IoT (internet of Things) is a combination of pervasive communication,

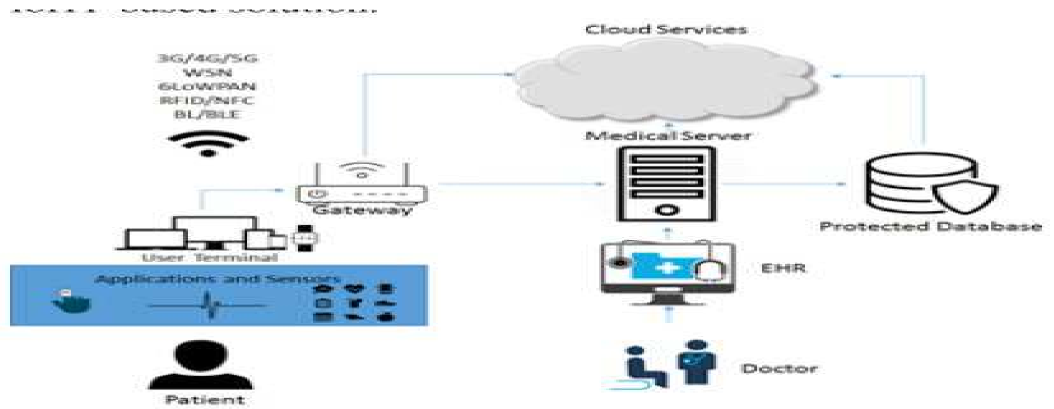


Fig. 4: Illustration of an IoMT-based solution architecture. IoMT [45]

Patients' data can be acquired using sensors and analyzed using programs designed for user terminals like computers, smart phones, smart watches, or even a dedicated embedded device. Over the IEEE 802.15.4 standard, the user terminal is connected to a gateway using short-range communication protocols such as Bluetooth low energy (BLE), Bluetooth, or 6LoWPAN (IPv6 over Low Power Wireless Personal Area Networks). For data processing and storage, this gateway connects to a (clinical) server or cloud services. In the other hand, patients' data can be stored in a health information system using electronic health records and, when the patient visits a medical doctor, he/she can easily access the clinic history of the patient. Figure 1 presents an illustration of an IoMT-based solution [45].

Some Challenges in Implementing Healthcare-based IoT Systems

Here, we outline some of those challenges, which are discussed widely in [46].

A. Standardization

Standardization deals with maintaining the standard of devices utilized in the healthcare industry. In recent days, IoT devices are being manufactured in sizably voluminous numbers by sundry local manufacturers who are concentrating on incrementing quantity while the standard is being compromised.

B. Continuous Monitoring

Continuous monitoring deals with body connected sensors utilized for analyzing and accumulating the physiological condition of the patient for perpetually observing the health of the patient in case of an emergency. There are very few devices available in the market having precise results.

C. Data Protection

Data Protection deals with providing security to health data. This data is considered crucial since it contains the information about body vitals as well as personal information. Ergo, bulwarking this data is considered crucial since any hindrance to health data from the outside environment can have devastating results and lead to life threatening issues.

Review of Machine Learning Frameworks

Within the ML framework, there are four main types of algorithms as seen in figure 5 below: supervised, unsupervised semi-supervised and reinforcement. Most of the ML studies [47; 48;49] performed in the area of ECG analysis aimed at disease diagnosis or risk stratification use supervised learning techniques. The goal of supervised learning is the inference of a function or a score from labeled or annotated training data. The two main supervised techniques are classification and regression, which differ on having as outputs categorical and continuous variables, respectively. Classification of different heart rhythms is one of the most developed applications of ML to the ECG using supervised approaches. Some popular algorithms include logistic regression, support vector machines, artificial neural networks, and random forests as reviewed in [50]. In supervised learning, labels for the training data is provided and/or select features to feed the algorithm to learn, whereas unsupervised learning algorithm is applied on raw data and learns fully automatic.

Unsupervised learning aims at discovering hidden structures in datasets with no previous knowledge about reference outcomes or labels [50]. The most common unsupervised learning method is clustering, used for grouping data with a similar data structure. Another common task in unsupervised learning is dimensionality reduction where principal component analysis (PCA) is one of the most frequently used methods in traditional ECG analysis, projecting data onto its feature subspace [51]. A recent case of success has been the use of mathematical modeling based on the Hermite functions and clustering techniques to identify ECG phenotypic subgroups in hypertrophic cardiomyopathy [50].

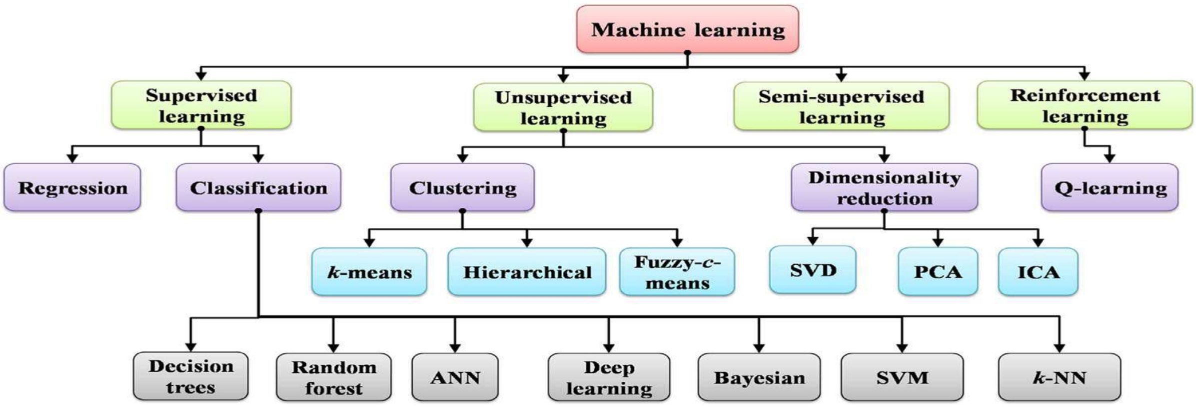


Fig. 5: Types of ML [51].

Most of the real-world application's data is the combination of labeled and unlabeled. The supervised learning algorithms work efficiently on the labeled information, and unsupervised learning works efficiently on unlabeled data [52].

Semi-supervised learning utilizes labeled and unlabeled data (mostly a smaller quantity of the former) in order to optimize the efficiency of classification. Semi-supervised learning is mainly useful when the human-power necessary to provide full labeling of all working data exceeds practical possibilities. However, it also conveys the advantage of keeping some distance from intrinsic biases provided by human operators, which probably exist even when such operators are considered experts. The semi-supervised learning introduced to work on the data with the combination of both labeled and unlabeled. It encompasses semi-supervised classification to perform classification on partially labeled data, constrained clustering to perform clustering with both labeled and unlabeled data, regression with unlabeled data and dimensionality reduction for labeled data [53].

There are two distinct goals of semi-supervised learning that are to predict the labels on unlabeled data in the training set and to predict the labels on future test data sets. Concerning these goals, semi-supervised learning divided into two categories: Transductive learning and inductive semi-supervised learning. Transductive learning is used to predict the exact labels for a given unlabeled dataset, whereas the inductive semi-supervised learning learns a function f: $X \mapsto Y$ so that f expected to be a good predictor on future data. Semi-supervised learning fits with several real-time applications such as natural language processing, classifying the web content, speech recognition, spam filtering, video surveillance, and protein sequence classification, etc [53]

Reinforcement learning (RL) algorithm continuously learns by interacting with the environment and gathers information to take certain actions. RL maximize the performance by determining the optimal result from the environment. Q-learning techniques is one of the model-free reinforcement learning approach [51]. Reinforcement learning is considered when there is a conjunction of other types of learning with the underpinning of constant trial and error. Reinforcement learning progresses without an initial notion of the objective and relies on constant interaction with new conditions and input in order to maximize the reward or benefit (performance), even if not immediately. Most notable examples of reinforcement learning can be found outside the medical field such as self-driving cars and advance systems for game implementations.

Table 1: Analysis of Different Types of Machine Learning Algorithms Applications with IoT

S/N	Year Of Work with Ref	Objective	Model Used	ML algorithm Used	Dataset Used	Application
1	[54].	Comparative Study of Classification Techniques to Predict the prevalence of heart disease.	Guided Learning classification Based	Support Vector Machine (SVM) with linear kernel/ Logistic Regression	Cleveland Heart Disease Dataset From UCI Repository	To minimize the human-based error chances in order to contribute to the medical science diagnosis and analysis.
2	[55].	Selecting	Guided	Linear	Random/Unknown	An approach

		Machine Learning Algorithms using Regression model	Learning regression Based	Regression Algorithm	dataset	to allot the ranking to the distinct machine learning algorithms in ranking list with by default setup of parameters using linear Regression.
3	[56]	Automation of an Intelligent Self Learning System for Home using IOT	Guided Learning classification	Naïve Bayes Algorithm	Dataset Generated Through Sensors	Provides the better assimilation of environment in home with lesser human interference and automatic fault detect ion in devices.
4	[57]	IOT Healthcare Analytics as importance of Anomaly Detection.	Unguided Learning	Anomaly Detection	Sensors	In Health Monitoring. To minimize the human-based errors in diagnosis, more reliable system, maximize the early detection rate of diseases.
5	[31]	IOT Service Clustering	Unguided Learning	MDM Algorithm,	Single and limited Size Sensors based	On the basis of clustering

		for Dynamic Service Matchmaking.		density-Peak-Based Clustering approach	Dataset	algorithm to dynamically discover, match making and exchange will be performed in efficient manner.
6.	[58]	Cluster Based Location Privacy in WSN for IOT	Unguided Learning-Cluster Based	K-Means	Sensor Generated	To provide K-means Cluster-based Location Privacy (KCLP) that enhanced the security and minimizes the delay in expense of energy consumption at minor level that can be reduced further in future work.
7	[59]	Traffic Data Classification for Security in IOT – Based Road Signaling System.	Guided Learning classification	Support Vector Machine (SVM)- at edge	Raw Traffic Dataset generated through 3 cities of London in time period of 5 yrs (2011-2016)	IOT Automated Traffic Signaling System in order to dynamically regulate the traffic in congested areas.

Table 2: Advantages and Disadvantages of Traditional ML Methods

Machine Learning Algorithms	Advantages	Disadvantages
The K-means method	Relatively efficient. Can process large data set. Intuitive algorithm with relative computational complexity.	• Often terminates at a local optimum • Applicable only when mean is defined • Not applicable for categorical data • Unable to handle noise data • Not suitable to discover clusters with non-convex/irregular shapes
Support Vector Machine (SVM)	• Can utilize predictive power of linear combinations of inputs • Good prediction in a variety of situations • Low generalization error • Easy to interpret results	• Weak in natural handling of mixed data types and computational scalability • Very black box • Sensitive to tuning parameters and kernel choice • Training an SVM on a large data set can be slow • Testing data should be near the training data
Decision Trees	• Some tolerance to correlated inputs • A single tree is highly interpretable • Can handle missing values • Able to handle both numerical and categorical data • Performs well with large datasets	• Cannot work on linear combinations of features • Relatively less predictive in many situations • Practical decision-tree learning algorithms cannot guarantee to return the globally-optimally decision tree • Decision-tree can lead to over fitting
Logistic Regression	• Provides model logistic probability • Easy to interpret • Provides confidence interval	• Does not handle the missing value of continuous variables • Suffers multicolinearlity • Sensitive to extreme values of continuous variables

	• Quickly update the classification model to incorporated new data	
Naïve Bayes	• Suitable for relatively small training data set • Can easily obtain the probability for a prediction • Relatively simple and straightforward to use • Can deal with some noisy and missing data • Can handle multiple classes	• Prone to bias when increasing the number of training data set • Assumes all features are independent and equally important, which is unlikely in real world cases • Sensitive to how the input data is prepared
Neural Networks	• Good prediction generally • Some tolerance to correlated inputs • Incorporating the predictive power of different combinations of inputs	• Not robust to outliers • Susceptible to irrelevant features • Difficult in dealing with big data with complex model

IoT based Healthcare Monitoring Using Machine Learning

Machine learning has revolutionized the healthcare industry where ML in integration with IOT is useful in determining the diagnosis and treatment along with remotely monitoring health status of patients. An Automated IOT based healthcare monitoring for Remotely located patients where implementation of this system helps doctors and guides them along with providing E-mails was proposed in [45] in case of abnormal conditions. This implemented system also provides an interface among doctors, nurses and concerned relatives of patients. This system's major purpose is to remotely monitor health status of a patient, simultaneously alert if heart rate goes in abnormal values and provides data analysis using supervised learning models such as SVM, KNN, Naive Bayes and J48 classifier. The system generated alert when result of decision support system shows the critical condition of patient.

In the system, remote patient healthcare monitoring was achieved by developing website and storing real time heart rate data on web server's database with IoT and provided prediction of heart disease with supervised machine learning classification techniques where J48 classifier performed better (92% accuracy) on the basis of predefined dataset. The probability of predicting heart problem will increase with generating rules from UCI heart database with decision tree algorithm. Thus, the system can use UCI heart rate dataset and provides an example of ML models in integration with IOT for healthcare monitoring [45].

In [60], a system for Chronic disease monitoring using IoT for vital sign monitoring using wearable devices such as heart rate, blood pressure and temperature sensor and the patient can determine these vital signs in the form of text document without the need of nurse. This system is trained for these vital sign data using data mining techniques (Support Vector Classifier, J48, and Naïve Bayes algorithms) and enables high risk patients to be timely checked in order to improve their quality of life. This system is primarily designed for elderly patients where raw data is collected from aforementioned sensors and model returns real-time data which can be healthy or unhealthy for particular patient. Main concentration of this system is towards problems related to health parameters such as Electrocardiogram (ECG), Oxygen Saturation (SpO2), Heart Rate (HR), Photoplethysmography (PPG), Blood Glucose (BG), Respiratory Rate (RR), and Blood Pressure (BP). For testing purpose, this system used instances of raw data from any patient.

For data processing, data mining approaches such as SVM, Naïve Bayes, J48 classifier are used where SVM performed better having an accuracy of 92%. The system captured the vital sign data for patients using medical sensor and noise from the raw data was removed. Afterwards, the raw data was used as test data in preprocessed step. Model training was learned by supervised methods (SVM, Naïve Bayes and J48). The model returned the vital sign data as healthy or unhealthy. From this status, patient takes medicine to detect early disease. The main drawback of the system was classification of large amount of data. Therefore, in that case Big data techniques and Map reduce techniques will be used to find health status as well as dependency can be found for further analysis.

[34] proposed a WONN-MLB ensemble for lung cancer disease in big data in 2019. The appropriate attributes were selected using an integrated Newton-Raphsons MLMR in the feature selection process to reduce the classification time. Later, the Boosted WONN Ensemble classification model was used to categorize the patient based on the selected characteristics, which increased accuracy and reduced FPR. [61] examined machine learning strategies for predicting surgical outcomes using TLE based solely on the structure of the brain connectome in 2015. In order to predict the outcomes, a two-stage connectome-based prediction framework was proposed, which significantly selected less irregular network connections to provide the outcomes of surgical therapy, and a linear kernel feature was used to improve the accuracy in every step.

For predicting surgery outcomes, [62] used machine learning techniques, particularly the accumulation of mutual data-based feature selection and supervised learning methods on multimodal information. As a result, MTLE gave the best anteromedial temporal lobectomy. The goal of Dai et al. (2015) was to reliably and efficiently predict heart-related hospitalizations on the basis of current medical history. Five machine learning techniques, such as SVM, AdaBoost, Logistic Regression, NB, and Likelihood Ratio Test differentiation, have been developed here. The training and testing datasets were used to train each system. In 2017, the machine learning techniques were well organized by [64] to effectively predict the outbreak of chronic illness in disease-frequent classes.

A system in [21] proposed a fast Fourier transformation-coupled machine learning ensemble technique for predicting short-term disease risk in chronic heart disease patients and providing effective clinical test recommendations. With an ensemble model, a combination of ANN, LS-SVM, and NB was developed. By clustering, eliminating noise, and using prediction techniques, [65] proposed an empirical strategy for predicting the disease. CART was used in order to create fuzzy rules. The new medical decision support model for the diagnosis of hepatitis based on RS and an ELM was recommended in 2013 by [66]. The proposed RS-ELM consisted of two phases; in the first step similar features were removed by the RS method from the dataset. ELM was subsequently classified using the other features.

[67] combined structural and series features in 2018 to predict HIV resistance by introducing SVM and RF techniques. On the basis of feature selection and extraction mechanisms, [68] intended to predict cancer via a hybrid model. In addition, to address the high-dimensionality problem, the established model used a filter feature selection technique called F-score and implemented an extraction method that used the mean methylation density,

symmetry between the methylation density and the mean methylation density, and FFT model of normal and cancer individuals for acceptable cancer classification and decreasing training time.

For improving the outcomes of sequence-based prediction models, [69] used a two-step method. The first step was focused on consensus learning, and the second step consists of the identification of the evaluated interactions based on the binding and network characteristics of genes in the gene regulatory network by SVM in both single and binary modes. [70] have used benefits of CNN for automated learning features from time series of needed signs and absolute feature embedding for heterogeneous clinical features to efficiently encode feature vectors. The features learned by CNN and statistical features by feature embedding were given to MLP to predict.

Healthcare Solution Model Based on Edge Computing and IOT

This focuses on how to intelligently use and prioritize network resources in a IOT framework over a secure and trust worthy transmission channel for a healthcare based application. This can be achieved by efficiently preprocessing the input data received from the sensor frameworks at the leaf devices. For this the leaf device, might take assistance of the back-end cloud servers that has access to heavy computing resources. The back end can perform the heavy number crunching and advise the end point device regarding the specific preprocessing to be done to be able to prioritize incoming data from sensor frameworks. The backend can use Machine Learning and data mining concepts to extract signatures from incoming sensor data and accordingly provide medical interpretation based on the captured data. Using this, the frontend device can provide an assessment of the patient health condition. The implementation can be extended to provide a method so that only prominent fluctuations can be provided to the back end where a physician can analyze the data and conclude. This will help in remote diagnosis and provide better rural medication where the doctor is away [71].

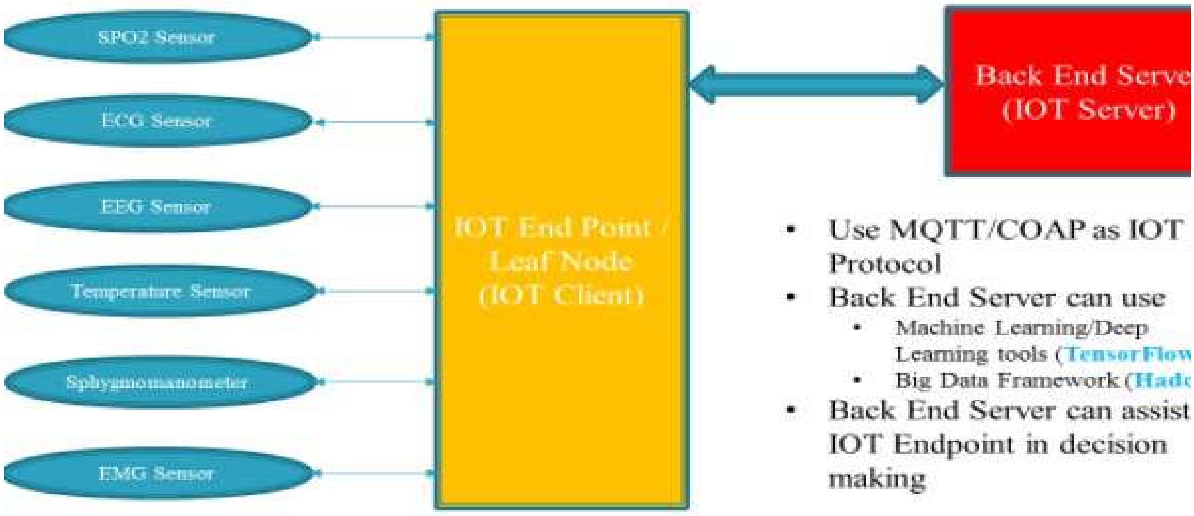


Fig. 6: **An IOT Endpoint.** [15]

As can be seen from fig. 6 above, An IoT Endpoint:
 Has multiple sensors connected to it.
 An IOT endpoint connected to a back-end server over a Wireless network
 The IOT Endpoint can communicate with the IOT Backend via IOT protocols (MQTT, CoAP)

The backend server

 Can perform computation intensive tasks as it has access to high end computing resources.
 For computation tasks, the backend can employ various techniques like big data analysis, Machine learning, neural networks
 Can be cloud based. However, our focus would be on edge computing
 Can send notifications to the IOT Endpoint

An MQTT client would be running on the IOT Endpoint and an MQTT server would be running in the cloud (Edge Server). This will provide the transport protocol necessary for communication between the IOT Endpoint and the data server.

The Hybrid Machine Learning Algorithms Proposed

In this work, the Support Vector Machine, Random Forest and the Naïve Bayes algorithms were used and finally hybridized. The choice for the above algorithms is due to their efficiency, sensitivity and accurate performance in their analysis and detection of heart diseases. From some literatures, we found that even with limited data samples, SVM can perform effectively. The SVM classifier is immune to the dimensionality curse since it can handle sparse data in high-dimensional dataset [72]. In addition, the SVM classifier offers superior generalization than the ANN and avoids the problem of local minima. The SVM approach works by generating the best separation plane under linearly separable conditions. Increasing the margin optimizes the hyperplane. The margin is the distance between each class's boundary and the nearest point. Support vectors are the spots closest to the boundary [73].

The following formulae help us to reach this aim.

Margin = $\frac{2}{||w||\text{^}2}$

W: normal vector of the hyperplane, the normal vector is perpendicular to the both hyperplanes which are parallel to each other, therefore, with the help of a normal vector we can calculate the distance between the two hyperplanes. In SVM, the aim is to choose w in a way that it can maximize the distance between hyperplanes [74].

Maximizing the margin is equivalent to minimizing the following formula:

L(W) = $\frac{||w||\text{^}2}{2}$

Moreover, an SVM could efficiently perform a nonlinear classification by using the kernel functions that map data into high-dimensional feature spaces. The radial kernel function used in this work takes the form:

K(xi, xi') = exp(-y + $\sum_{j=1}^{p}(xij - xi'j)^2$ y is a positive constant. If a given test observation x* = (x*1--- x*p), T is far removed from a training observation xi in terms of the Euclidean distance, then $\sum_{j=1}^{p}(xij - xi'j)^2$ will be large, and so k(x*, x_i) = exp(- $\sum_{j=1}^{p}(xij - xi'j)^2$) = will be very small. This means that in the nonlinear function

f (x) = $x\beta 0 + \sum \alpha ik(x, xi)$

x_i will have no effect on f(x*). The predicted class label for X* (test observation) is based on the sign of f(x*), so the training observations that are far from x* will not have any effect on predicting the class label for x*. In other words, the radial kernel exhibits local behavior because merely nearby training observations play a role in the predicted class label for a test observation.

Naïve Bayes Model

The naïve bayes model is based on Bayes' theorem through independent features. This model is trained to make it possible to calculate the posterior probability based on prior probabilities where every instance of D is allotted to the class of highest subsequent probability [75]. The basic idea of a Naïve Bayes classifier is to measure the posterior probability based on the prior probability and the likelihood. Given a dataset *X* with *[X1, X2, ... , Xn]* which represent *n* symptoms or the physiological conditions, the probability *P(Disease|X)* is the (posterior) probability of the patient having a disease and *P(NoDisease|X)* be the probability of the patient not having the disease. Given the set of symptoms and physiological conditions, this probability can be found using,

$$P(Disease|X) = \frac{\mathrm{P(X|Disease)} * \mathrm{P(Disease)}}{P(X)}$$

Or $P(Disease|X) \propto \mathrm{P(X|Disease)} * \mathrm{P(Disease)}$

But, in Naïve Bayes Classifier, every symptom is considered independent of the other. No symptom is considered an effect of any other symptom. Hence, the above equation becomes,

$P(Disease|X) \propto \prod_{i=1}^{n} P(Xi)P(NoDisease)$

Also, if probability of having the disease is only 50%, i.e., $P(Disease) = P(NoDisease)$ then, the posterior probabilities depend only on the prior probabilities, i.e., $P(Disease|X) \propto \prod_{i=1}^{n} P(Xi)$.

The decision on whether a patient has a given disease can be found using the maximum of the posterior probabilities.

If $P(Disease) < P(NoDisease)$

The prediction is made that the patient does not have the disease; otherwise, the prediction is made that the patient has the disease. The effectiveness of the Naïve Bayes Classifier depends on the quality of the training examples provided.

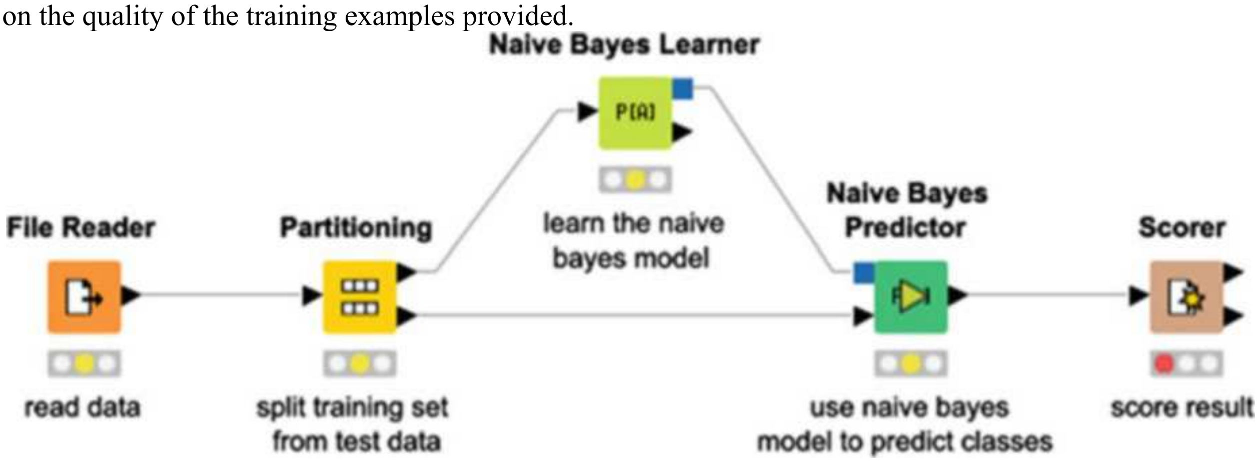


Fig. 10: The flow diagram of Naive Bayes in machine learning. Source: [76]

The diagram above in fig. 10 shows how data flow in Naïve Bayes model of machine learning algorithm.

Random Forest Model

Random Forest is one of the most popular and most powerful machine learning algorithms. It is a type of ensemble machine learning algorithm called Bootstrap Aggregation or bagging [77]. The random forest (RF) is a hierarchical collection of tree structured base classifiers. Text data usually have much number of dimensions. The dataset contains a large number of irrelevant attributes. Only few important attributes are informative for classifier model. RF algorithm uses a simple predetermined probability to select the most important relevant attribute. Breiman formulated the RF algorithm using sample data subsets and to construct multiple decision trees by mapping random sample of feature subspaces. The RF algorithm associated with a set of training documents D and Nf features can be described as follows:

(1) Initial: $D1$, $D2$,…….DK sampled by predetermined probability with replacement.

(2) For each document DK construct a decision tree model. The training documents are randomly sampled using its subspace of m-try dimension from the available features. Calculate all possible probability based on the m-try features. The leaf node produces the best data split. The process will be continued till it reaches the saturation criterion.

Combine the K number of unpruned trees $h1(X1)$, $h2(X2)$, into a random forest ensemble and use the high probability value for classification decision [73].

Analysis of Existing System

The present systems analyzed lacked the accuracy and efficiency needed for a smart health data management. The systems did not use hybridized parameters and machine learning models to implement their findings, thus presenting a loophole. We also found out that most systems only used internet of things only to detect cardiovascular diseases. Some used machine learning to actualize their goals but there was no system that conceptualized the hybridization of both technologies to enhance accuracy, productivity and efficiency in data management.

A review on the research work by [54], they designed an IOT based MPM (Multi-parameter Patient Monitor) system where four parameter namely heart rate, respiration rate, oxygen saturation and temperature are monitored using corresponding sensors and an email is sent to patient's guardian in case of emergency. The project also focuses on improving the performance of MPM system using Support Vector Machine (SVM) algorithm. The classification accuracy of 95% has been achieved. The block diagram of the existing system is as below:

.

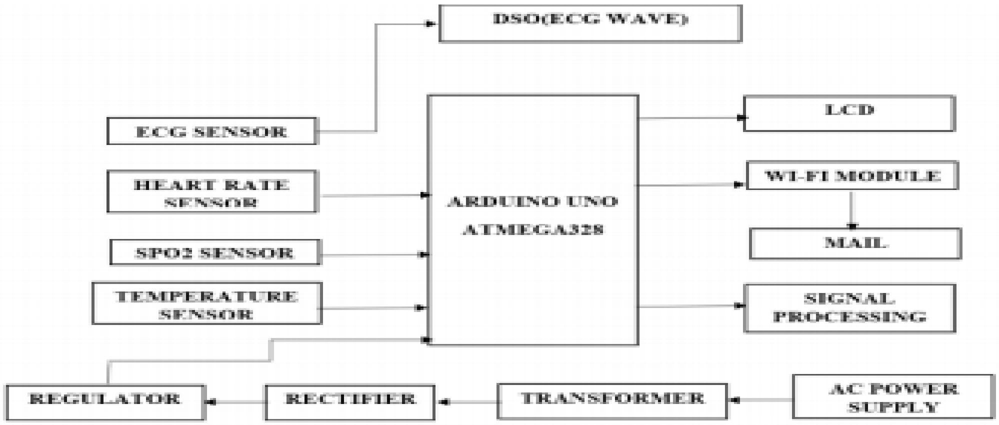


Fig. 7: Block Diagram of Existing System

Analysis of the Proposed System

The proposed system has been analyzed using a block diagram and OOP.

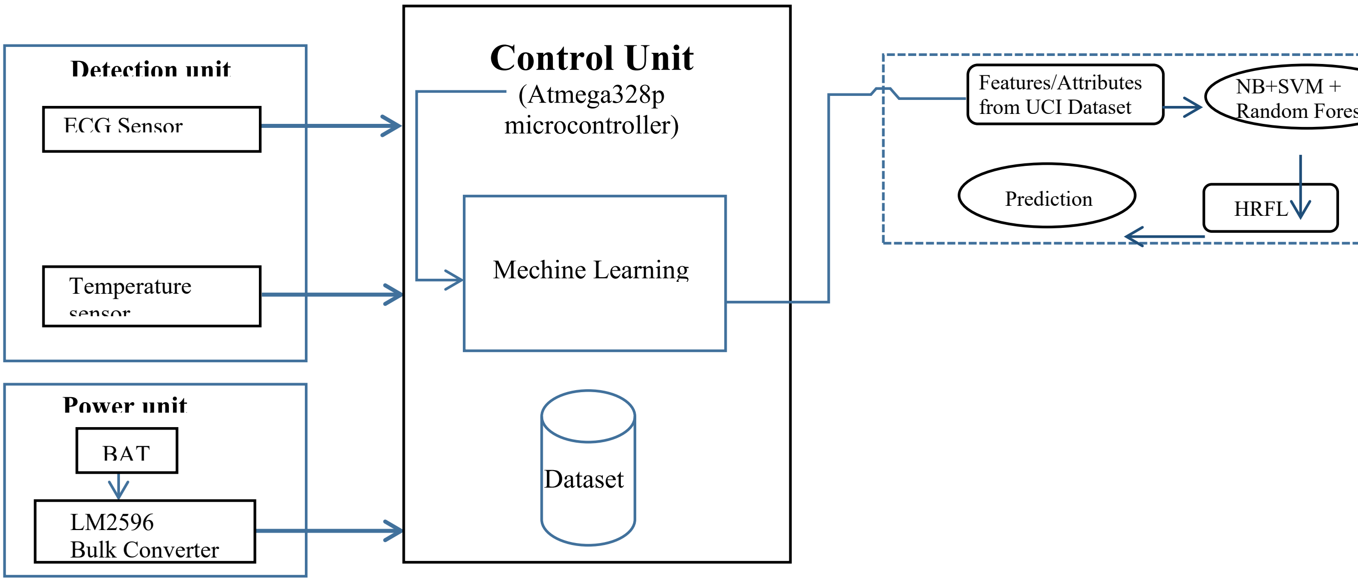


Fig. 8: Block diagram of the Proposed System

The proposed system is made up of the control, detection and feedback units. The detection units are the ECG Monitor and Temperature sensor. The ECG Monitor reads heart rate data from the human body and gives analog output. The temperature sensor reads the temperature of the human body and gives digital output to the microcontroller which is the main control unit. The microcontroller receives sensor signals or outputs and sends them to machine learning sub control unit for necessary analysis, predictions and data visualization. The inclusion of machine learning sub control unit for sensor data analysis before the general prediction is made is one of the frameworks that distinguishes the existing system from the proposed system.

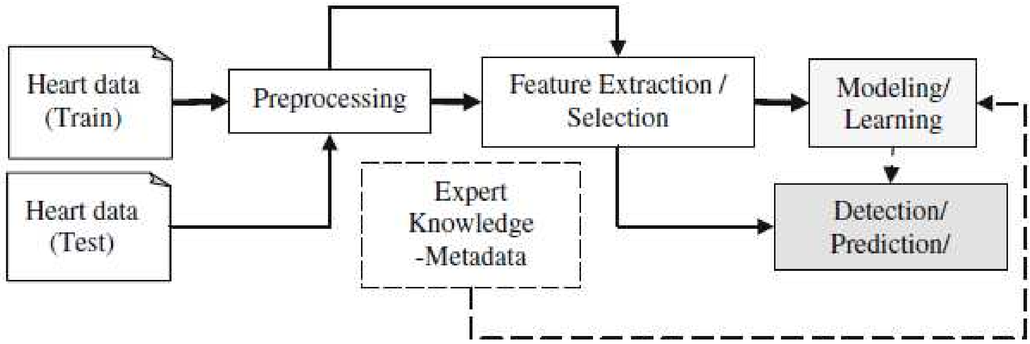


Fig. 9: **Data Training/Testing System Architecture**

The figure 14 above shows how we train and test data before prediction. We trained our model with 75 percent of the dataset and 25 percent was used for the testing of the model.

Methodology

In this study, hybrid machine learning algorithms are chosen to detect the presence or to decide the probability of having heart diseases from a large dataset with dataset generated from sensors. The algorithms used are Support Vector Machine (SVM), Naïve Bayes Classifier and Random Forest. Then a continuous disease monitoring system designed to interface Python language with Arduino based microcontroller system and other sensors. Object oriented Analysis and Design (OOAD) is used in the design and development of the proposed system. The system design concepts and components are presented using the Python language for data analysis and prediction, Unified Modeling Language (UML), C++ and some other interactive programming languages are used for both the development of the machine learning and IoT components

System Design and Implementation

The data set used in this research for training and validating the model is gotten from Cleveland heart disease dataset from UCI machine learning repository with the parameters as shown in figure 25 below.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0
5	58	0	0	100	248	0	0	122	0	1.0	1	0	2	1
6	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
7	55	1	0	160	289	0	0	145	1	0.8	1	1	3	0
8	46	1	0	120	249	0	0	144	0	0.8	2	0	3	0
9	54	1	0	122	286	0	0	116	1	3.2	1	2	2	0

Fig. 11: **Dataset Preview**

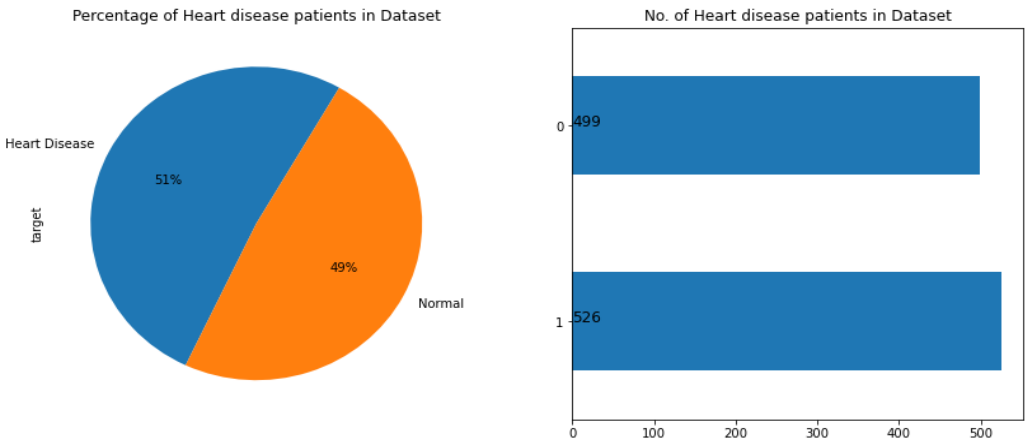


Fig 12: Patient with and without Heart Disease

The above fig. 12 shows that the dataset used in training the models contain 526 patients with heart disease and 499 patients without heart disease.

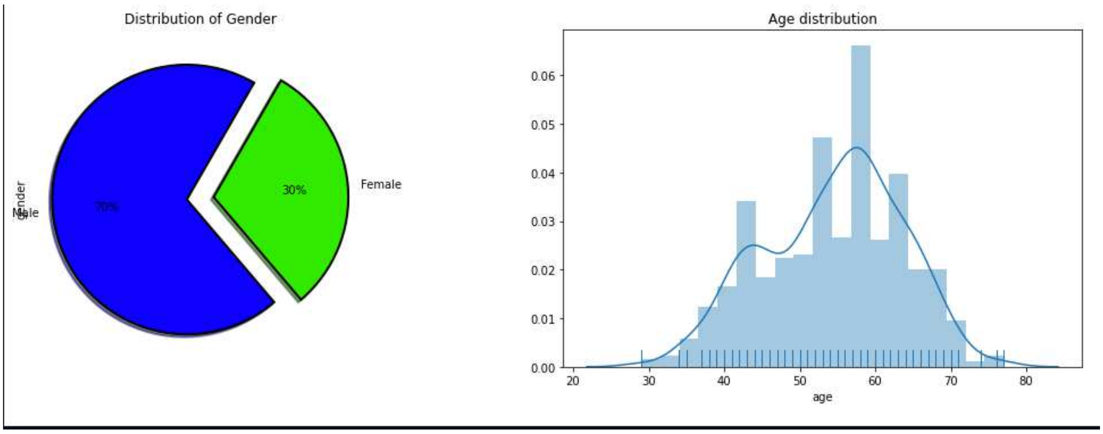


Fig. 13: **Gender and age distribution**

The gender distribution chart in fig 27 shows that the data set contain 70% men and 30% female while the age distribution chart shows that the age ranges from (29-78)

Age and Gender Distribution of Normal Patients

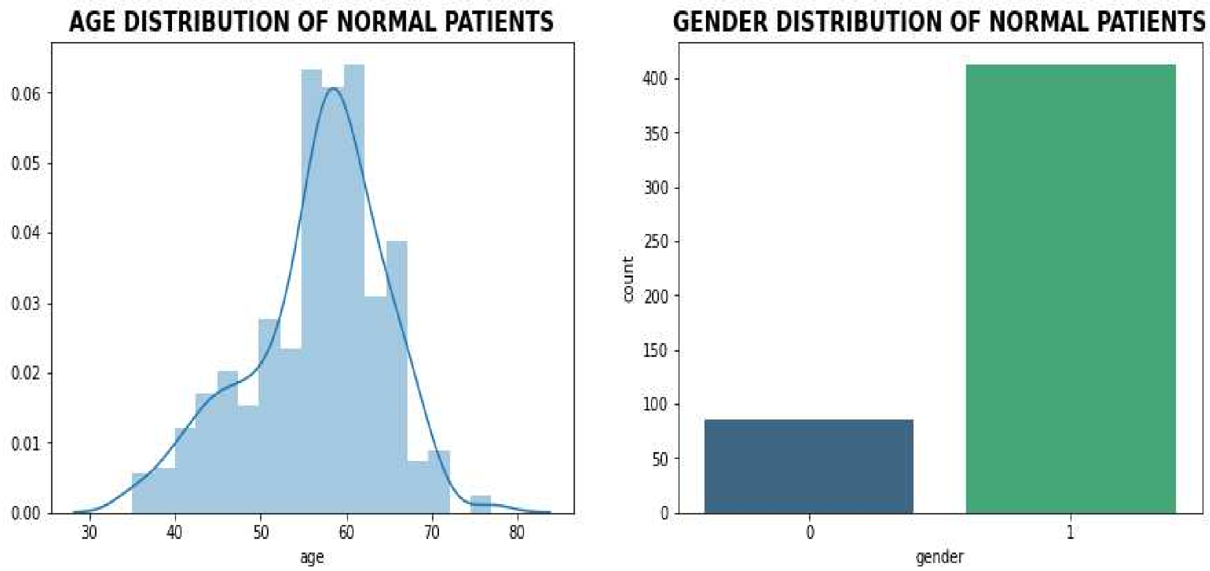


Fig 14: **Age and Gender Distribution of Normal Patients**

The age distribution chart in fig 28 shows the age range of normal patients in the data set while gender distribution chart shows that more men were not suffering from heart disease.

Age and Gender Distribution of Patient with Heart Disease

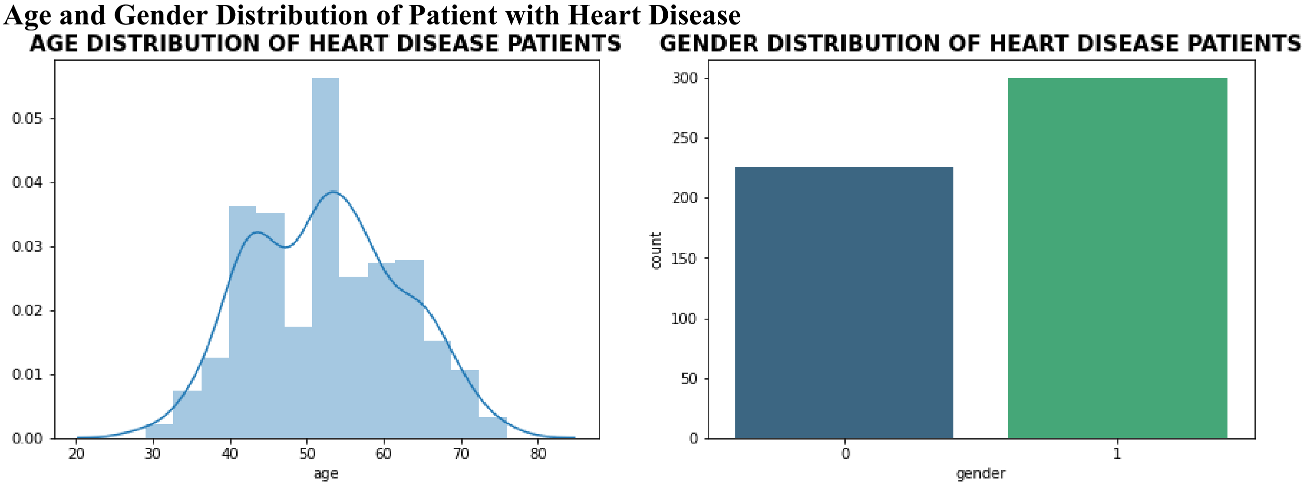


Fig. 15: **Age Distribution and Gender Distribution of Patient with Heart Disease**

As shown in fig. 15 above, patients between the ages of 50-55 suffer from heart disease more. Also, in gender distribution chart, more men suffer from heart disease.

Pair Plot

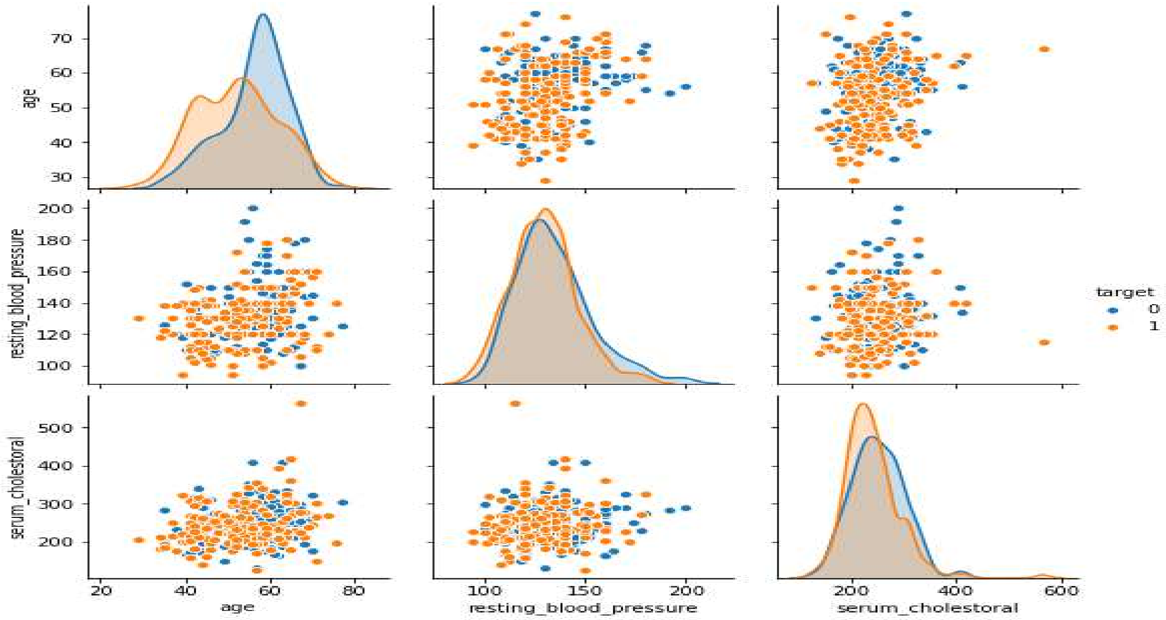


Fig 17: **Pair Plot of the some features in the data set**

Fig. 17 shows the relationship between age, resting_blood_pressure and serum_cholestoral. Also, the diagonal subplots are the univariate histograms (distributions) for each attribute.

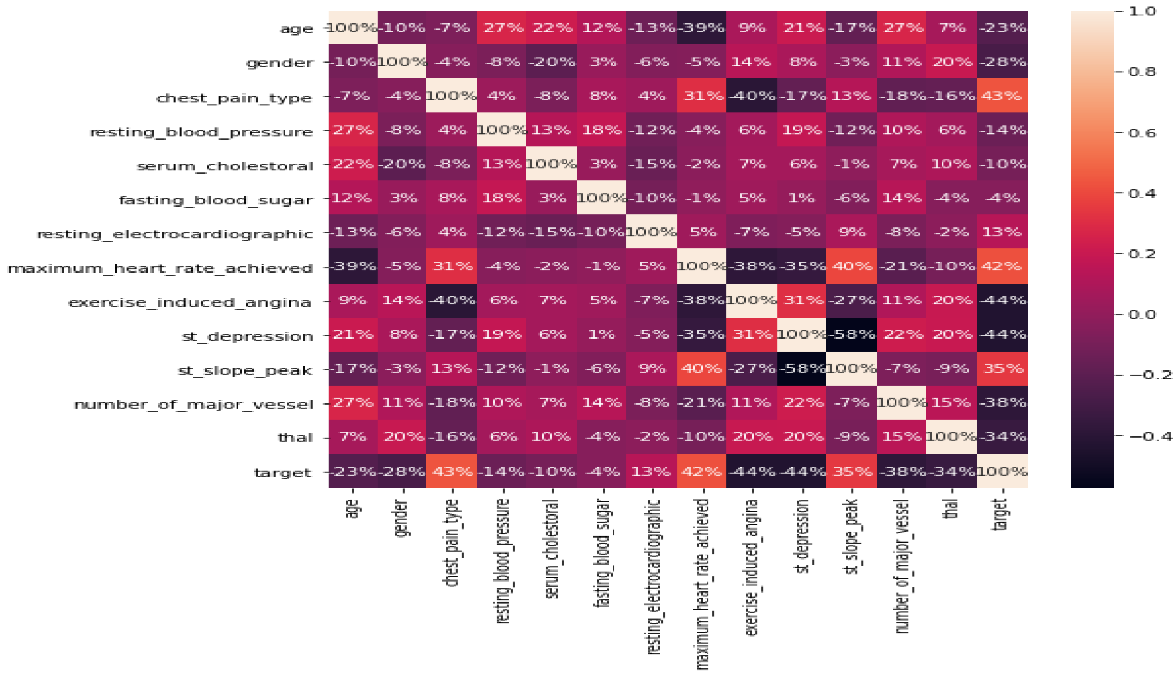


Fig 18: **Correlation Table**

Fig. 18 above shows the correlation of attribute in the data set. As we can see, maximum_heart_rate_achieved has a correlation of 30% with chest_pain_type.

Precision and Recall for the Models

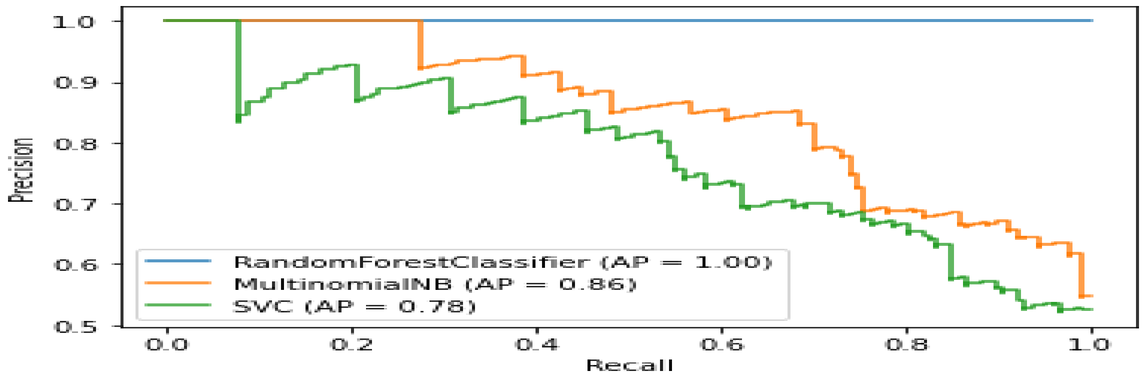


Fig. 19: **Precision and Recall for Evaluating the models**

In fig. 19 above, the precision and recall of random forest model exceed 100%. 75% precision and 65% recall for Naïve Bayes model. 77% precision and 56% recall for support vector machine. From the chart in fig 1.8, random forest performed better than Naïve Bayes and support vector machine.

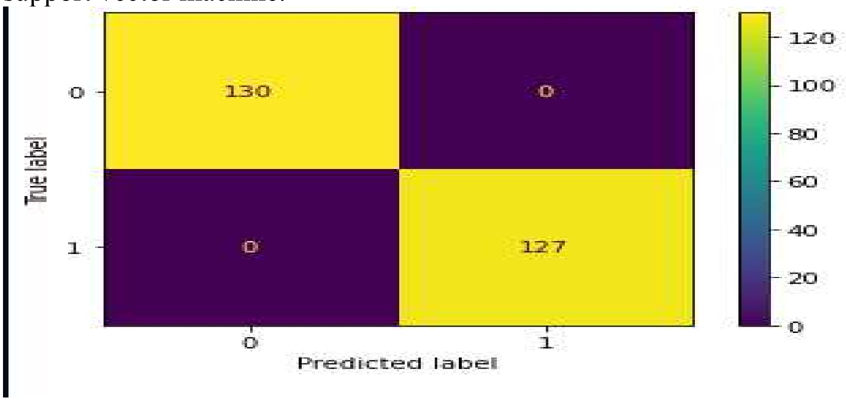


Fig. 20: **Confusion Matrix for Random Forest Model**

As shown in fig. 20 above, the four square box correspond to True positive (TP), True negative (TN), False positive (FP) and False negative (FN). From the figure above, we see that the model was able to learn without any confusion since TN and FN are zero.

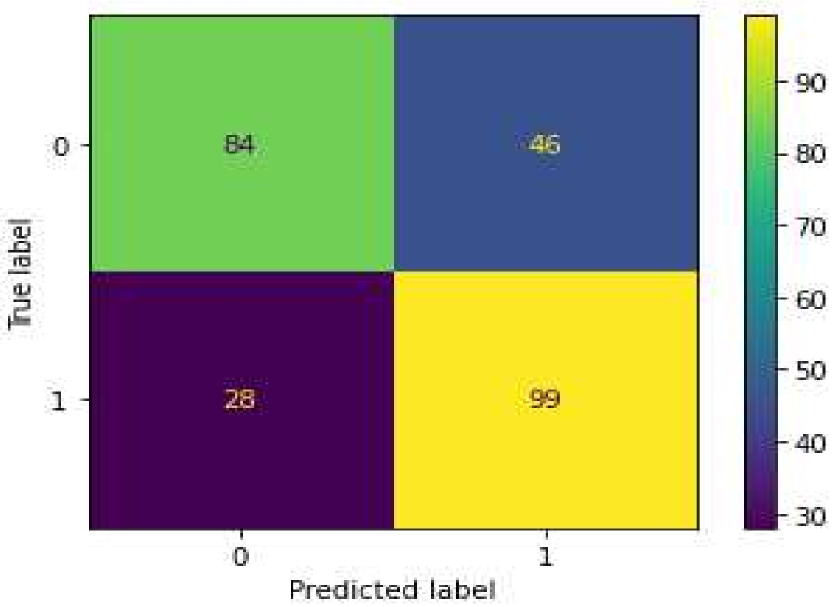


Figure 34: **Confusion Matrix for Naïve Bayes Model**
In figure 34 above, the Naïve bayes model misclassified 28 patients with no heart disease as having heart disease and 46 patients with heart disease as not having heart disease.

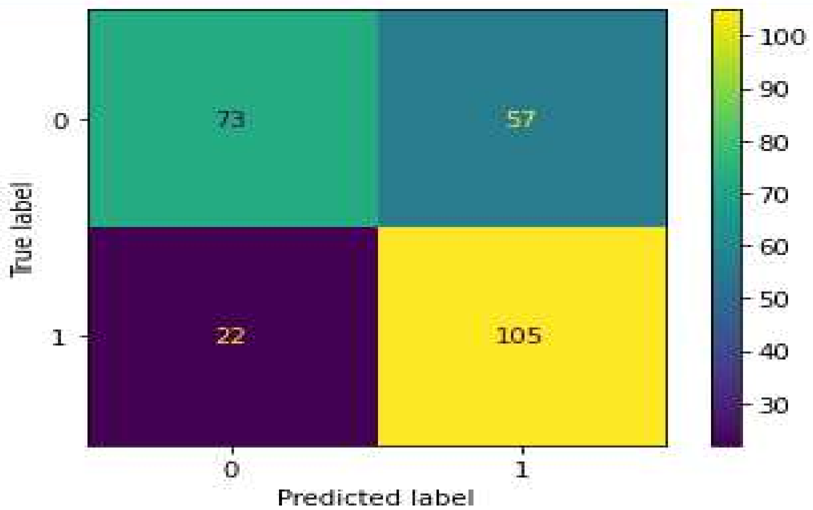


Fig. 35: **Confusion Matrix of Support Vector Machine**
In fig. 35 above, the SVM model misclassify 22 patients with no heart disease as having heart disease and 57 patients with heart disease as not having heart disease. This makes the accuracy of the model to reduce.
The confusion matrix is helpful to predict the classification problems. In the predicted class, the total number of exact predictions for a class goes into the expected row for that class value. In the same manner, the total number of incorrect class predictions goes into the expected row for that class value and the class value of predicted column.
When applying sample test data of each disease to its trained model, the model trained with random forest algorithm gives accurate results when compared to the models trained with

Naive Bayes and Support Vector Machine algorithms. The main reason for the performance of random forest algorithm against test data is the property of high-dimensional feature and self-judges the essential features in the dataset. Feature interaction is also recognized by the random forest algorithm.

4.7 Receiver Operating Characteristic Curve

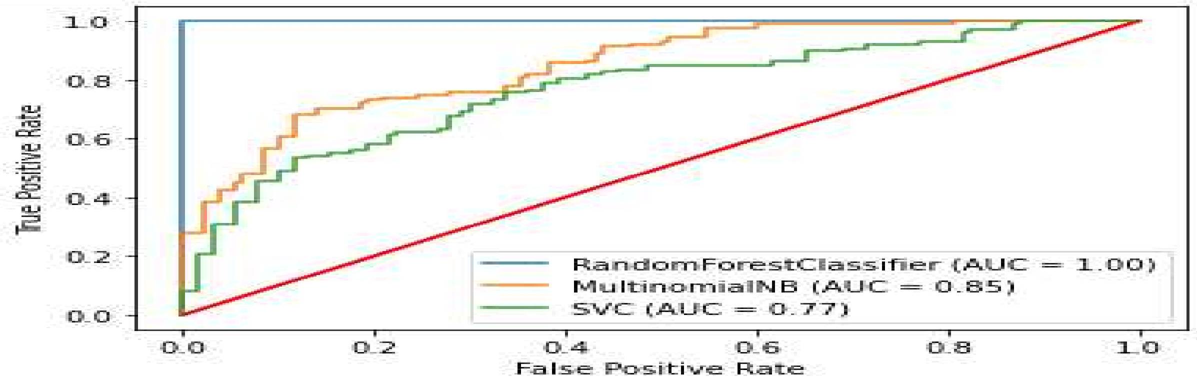


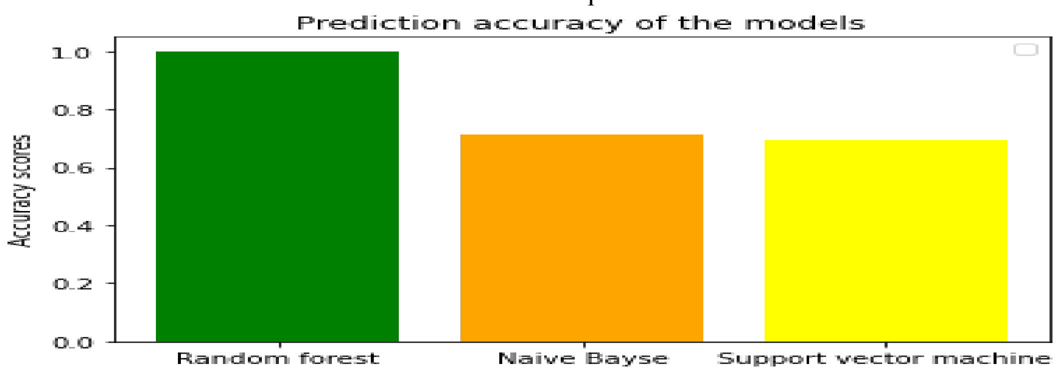
Fig. 36: **Receiver Operating Characteristic Curve (ROC_curve)**

ROC_curve shows the performance of a classification model at all classification thresholds. As shown in fig. 36 above, the ROC_curve plots two parameters;

True Positive Rate (TPR) = TP/TP+FN

False Positive Rate (FPR) = FP/FP+TN

The ROC (receiver operating characteristic) curve is a promising tool used to predict the probability of a binary outcome. The graph is plotted for a number of different candidate threshold values between 0.0 and 1.0 while keeping the false positive rate on the x-axis and the true positive rate on the y-axis. The prime utilization of the ROC curve is in deciding where to draw the line between “normal” and “not normal.” The decision will become easier if all the control values exceed or are lower than all the patient values.



Fig, 37: **Accuracy of the Models**

Summary

In this work, early identification and processing of raw heart data helped in the long term saving of human lives and early detection of abnormalities in heart conditions. The ECG and temperature sensors with support vector machine, random forest and naïve bayes machine learning models were used in this work to analyze and process raw data from sensors and provide a new and novel discernment towards heart disease. The mortality rate was drastically controlled as the heart disease is detected at the early stages and preventative measures adopted as soon as possible. The combined approach of using both IoT and machine learning algorithms for patient monitoring illustrated. The IoT component, in addition to the analyzed dataset, is mainly involved in sensing the parameters and communicating it to a monitoring device. The monitoring device can use a machine learning approach to respond back or can intimate a doctor in case of an emergency. The work specifies how Random Forest, Support Vector Machine and Naïve Bayes Classifiers, as hybrid, were used in predicting the heart disease based on some attributes.

The dataset is chosen from online repositories. The techniques of pre-processing applied are filled in missing values and removing correlated columns. Next, the classifiers are applied to the preprocessed dataset, and then Support Vector Machine, Naïve Bayes and random forest models are constructed. The smartness of this work is the analysis made by the machine learning models in order to accurately predict heart diseases in patients. This work is an IoT based Machine learning that processes sensor data and identify the most significant clinical parameters to get heart disease, send the sensed data to a hybrid machine learning model for data analysis and prediction. The most significant clinical parameters that indicate impending heart diseases are identified with the help of ROC analysis.

References

[1] Sharma, K. and Nandal, R. (2019) "A Literature Study On Machine Learning Fusion With IOT," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, pp. 1440-1445

[2] Bhadoria, R. K., Saha, J., Biswas, S., & Chowdhury, C. (2021). IoT-based location-aware smart healthcare framework with user mobility support in normal and emergency scenario: a comprehensive survey. In *Healthcare Paradigms in the Internet of Things Ecosystem*. Elsevier Inc. https://doi.org/10.1016/b978-0-12-819664-9.00007-7

[3] Bansal, M., & Gandhi, B. (2019). *IoT & Big Data in Smart Healthcare (ECG Monitoring). 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).*

[4] Hussain, F., Hussain, R., Hassan, S. A., & Hossain, E. (2020). Machine Learning in IoT Security: Current Solutions and Future Challenges. *IEEE Communications Surveys and Tutorials*, *22*(3), 1686–1721. https://doi.org/10.1109/COMST.2020.2986444

[5] Fernandes C. O. and Lucena C. J. P.(2017). A software framework for remote patient monitoring by using multi-agent systems. JMIR Med Inform;5(1):e9.

[6] Dharmasiri, N. D. K. G., & Vasanthapriyan, S. (2018). *Approach to Heart Diseases Diagnosis and Monitoring through Machine Learning and iOS Mobile Application. 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)*

[7] Hantao, H. and Hao, Y. (2019). Compact and Fast Machine Learning Accelerator for IoT Devices; Computer Architecture and Design Methodologies; Springer: Singapore, pp. 9– 11

[8] Tyagi, S., Agarwal, A., & Maheshwari, P. (2016). A conceptual framework for IoT-based healthcare system using cloud computing. *2016 6th International Conference – Cloud System and Big Data Engineering (Confluence).* doi:10.1109/confluence.2016.7508172

[9] Al-milli, N (2013) "Backpropogation neural network for prediction of heart disease", *Journal of Theoretical Applied Information Technology*, 56(1), pp. 131-135.

[10] Baccour, L (2018). "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets", *Expert Syst. Appl.*, 99(10), 115-125, doi: 10.1016/j.eswa.2018.01.025

[11] Dineshkumar Jaiswar, Sanjna S. Repal (2015) "Real Time Tracking and Health Monitoring of Soldiers using ZigBee Technology: a Survey" *International Journal of Innovative Research in Science, Engineering and Technology*, **4(7)**.

[12] Hassan, M. S., Rahman, A., & Reza, A. W. (2019). *Health status from your body to the cloud: The behavioral relationship between IoT and classification techniques in abnormal situations. Sensors for Health Monitoring, 135–156*

[13] Ross, E.G., Shah, N.H., Dalman, R.L., Nead, K., Cooke, J., Leeper, N.J.,(2016). The use of machine learning for the identification of peripheral artery disease and future mortality risk. J. Vasc. Surg. 64, 1515_1522

[14] Ganesan, M., & Sivakumar, N. (2019). IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*

[15] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). *Effective Heart Disease Prediction using Hybrid Machine Learning Techniques. IEEE Access, 1–1*

[16] Nashif, S., Raihan, R., Islam, R. and Imam, M.H. (2018) Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System. *World Journal of Engineering and Technology*, 6,854-873.

[17] Das, R., Turkoglu, I. and Sengur, A.(2009). "Effective diagnosis of heart disease through neural networks ensembles," *Expert Syst. Appl.*, 36(4), pp. 7675-7680, doi: 10.1016/j.eswa.2008.09.013.

[18] Nahar, J., Imam, T., Tickle, K. S and Chen, Y. P. P (2013). "Association rule mining to detect factors which contribute to heart disease in males and females," *Expert Syst. Appl.*, 40(**4**) pp. 1086-1093, 2013. doi: 10.1016/j.eswa.2012.08.028

[19] Zaman, S and Toufiq, R (2017). "Codon based back propagation neural network approach to classify hypertension gene sequences," in *Proc. Int. Conf. Elect., Comput. Commun. Eng. (ECCE)*, pp. 443-446

[20] Ravish, D. K., Shanthi, K. J., Shenoy, N. R and Nisargh, S. (2014). "Heart function monitoring, prediction and prevention of heart attacks: Using artificial neural networks," in *Proc. Int. Conf. Contemp. Comput. Inform. (IC3I)*, pp. 1-6.

[21] Zhang, W. and Han, T. (2017). "Towards heart sound classification without segmentation using convolutional neural network," in *Proc. Comput. Cardiol. (CinC)*, vol. 44, pp. 1-4.

[22] Banu N. K. S. and Swamy, S. (2016). "Prediction of heart disease at early stage using data mining and big data analytics: A survey," in *Proc. Int. Conf. Elect., Electron., Commun., Comput. Optim. Techn. (ICEECCOT)*, pp. 256-261.

[23] Saini, R., Bindal, S. and Bansal, P (2015). "Classification of heart diseases from ECG signals using wavelet transform and kNN classifier," *Proc. Int. Conf. Comput. Commun.*, pp. 1208-1215, doi: 10.1109/CCAA.2015.7148561.

[24] Sanket R. Jadhav , S. S. Bhabad, (2019). Heart Rate Monitor using IOT, International Journal of Engineering Research & Technology (IJERT) Volume 08, Issue 06

[25] Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). *Machine Learning in Healthcare: A Review. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*

[26] Meidan, Y., Bohadana, M., Shabtai, A., Guarnizo, J. D., Ochoa, M., Tippenhauer, N. O. and Elovici, Y. (2017). "Pro_lIoT: A machine learning approach for IoT device identification based on network traffic analysis," in *Proc. Symp. Appl. Comput.*, pp. 506-509

[27] Wu, J. Luo, S., Wang, S., and Wang, H. (2019). "NLES: A novel lifetime extension scheme for safety-critical cyber-physical systems using SDN and NFV," *IEEE Internet of Things J.*, 6(2), pp. 2463-2475.

[28] Li, G., Wu, J., Li, J., Wang, K. and Ye, T. (2018). "Service popularity-based smart resources partitioning for fog computing-enabled industrial Internet of things," *IEEE Trans. Ind. Informat.*, 14(**10**), pp. 4702-4711.

[29] Wu, J. Ota, K., Dong, M and Li, C. (2016). "A hierarchical security framework for defending against sophisticated attacks on wireless sensor networks in smart cities," *IEEE Access*, vol.

4, pp. 416-424
[30] Li, H. Ota, K. and Dong, M. (2018). "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE Netw.*, 32(**1**), pp. 96-101.
[31] Li, S., Xu, L., & Zhao, S. (2018). 5G internet of things: A survey. *Journal of Industrial Information Integration,* 10, 1–9.
[32] Wu, J. Ota, K., Dong, M and Li, C. (2016). "A hierarchical security framework for defending against sophisticated attacks on wireless sensor networks in smart cities," *IEEE Access*, vol. 4, pp. 416-424
[33] Li, H. Ota, K. and Dong, M. (2018). "Learning IoT in edge: Deep learning for the Internet of Things with edge computing," *IEEE Netw.*, 32(**1**), pp. 96-101.
[34] Zhang Z. (2015). Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction, *IEEE Trans. Biomed. Eng*. 62 (8)1902–1910.
[35] Jindal V. (2016). Integrating mobile and cloud for PPG signal selection to monitor heart rate during intensive physical exercise, in: Proceedings of the International Conference on Mobile Software Engineering and Systems, MOBILESoft '16, ACM, New York, NY, USA, pp. 36–37,
[36] Ukil A., Bandyoapdhyay S., Puri C., Pal A., and Mandana K. (2016). Cardiac condition monitoring through photoplethysmogram signal denoising using wearables: can we detect coronary artery disease with higher performance efficacy? Computing in Cardiology Conference ComputinginCardiology,pp.1–4
[37] Elsaadany Y., Majumder A.J.A, and Ucci D.R. (2017). A wireless early prediction system of cardiac arrest through IoT, Proc. Int. Comput. Softw. Appl. Conf. 2 (2017) 690–695,
[38] Chen, M., Zhang, Y., Qiu, M., Guizani, N. and Hao, Y. (2018). "Spha: Smart personal health advisor based on deep analytics", IEEE Commun. Mag. 56 **(3)**, pp. 164–169.
[39] Pozo, F. Arruga, I., Mujica, L.E, Ruiz, M. and Podivilova, E. (2016) Detection of structural changes through principal component analysis and multivariate statistical inference, Struct. Health Monit. 15(2), pp. 127–142.
[40] Yao, C. Qu, Y., Jin, B., Guo, L., Li, C., Cui, W. and Feng, L. (2016). "A convolutional neural network model for online medical guidance", *IEEE Access* 4, pp. 4094–4103.
[41] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A. and Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network, *IEEE Trans. Med. imaging* 35(**5**), pp. 1207–1216.
[42] Savran, A., Cao, H., Nenkova, A. and Verma, R. (2015). Temporal bayesian fusion for affect sensing: combining video, audio, and lexical modalities, *IEEE Trans. Cybern*. 45(**9**), pp. 1927–1941.
[43] Chen, M., Zhang, Y., Qiu, M., Guizani, N. and Hao, Y. (2018). "Spha: Smart personal health advisor based on deep analytics", IEEE Commun. Mag. 56 **(3)**, pp. 164–169.
[44] Chawla, M. N. (2020). AI, IOT and Wearable Technology for Smart Healthcare-A Review. *International Journal of Recent Research Aspects*, *7*(March), 9–13.
[45] Rodrigues, J. J. P. C., De Rezende Segundo, D. B., Junqueira, H. A., Sabino, M. H., Prince, R. M. I., Al-Muhtadi, J., & De Albuquerque, V. H. C. (2018). Enabling Technologies for the Internet of Health Things. *IEEE Access*, *6*(1), 13129–13141. https://doi.org/10.1109/ACCESS.2017.2789329
[46] Rose, K., Chaplin L. and E. S. (2015) "The Internet Of Things: An Overview: Understanding the Issues and Challenges of a More Connected World," [Online]. Available: https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-IoT-Overview-20151221-en.pdf. [Accessed 25 Feb 2021].
[47] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., Sun, R., & Garciá-Magarinõ, I. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, *2018*. https://doi.org/10.1155/2018/3860146
[48] Khairuddin, A. M., Ku Azir, K. N. F., & Kan, P. E. (2017). Limitations and future of electrocardiography devices: A review and the perspective from the Internet of Things. *International Conference on Research and Innovation in Information Systems,*

ICRIIS. https://doi.org/10.1109/ICRIIS.2017.8002506

[49] Jain, D. (2021). Industrial IoT. In *Indian Journal of Computer Science* (Vol. 6, Issue 1). https://doi.org/10.17010/ijcs/2021/v6/i1/158188

[50] Lyon A, Mincholé A, Martínez JP, Laguna P and Rodriguez B. (2018). Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances. J R Soc Interface ;15

[51] Lyon A, Ariga R, Mincholé A, Mahmod M, Ormondroyd E, Laguna P, et al.(2018). Distinct ECG phenotypes identified in hypertrophic cardiomyopathy using machine learning associate with arrhythmic risk markers. Front Physiol.;9:213

[51] Praveen K., Tarachand A. and Annavarapu C. S. R. (2019). Machine learning algorithms for wireless sensor networks: A survey. Information Fusion, 49:1–25.

[52] Kirti S. and Rainu N (2019). *A Literature Study On Machine Learning Fusion With IOT* Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019).

[53] Yoo J., Kim W. and Kim H. J. (2015). Distributed estimation using online semi-supervised particle filter for mobile sensor networks, IET Control Theory Appl. 9 (3)418–427

[54] Khanna, Divyansh, *et al*. (2015). "Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease." *International Journal of Machine Learning and Computing.* Vol. 5, Issue 5, pp 414

[55] Doan, Tri, and Jugal Kalita(2015). "Select ing machine learning algorithms using regression models." *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE.

[56] Bhide, Vishwajeet Hari, and Sanjeev Wagh (2015). "i-learning IoT: An intelligent self learning system for home automat ion using IoT." *2015 International Conference on Communications and Signal Processing (ICCSP)*. IEEE

[57] Ukil, S. Bandyoapdhyay, C. Puri and Pal A. (2016). "IoT Healthcare Analytics: The Importance of Anomaly Detect ion". *IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Crans-Montana. pp 994-997

[58] Han, Guangjie, (2018). "KCLP: A k-Means Cluster-Based Location Privacy Protect ion Scheme in WSNs for IoT." *IEEE Wireless Communications,* 2018. Vol. 25, Issue 6, pp 84 -90.

[59] Mookherji, Srijanee, and Suresh Sankaranarayanan (2019). "Traffic Data Classification for Security in IoT-Based Road Signaling System". *Soft Computing in Data Analytics*. Springer, Singapore, 2019. pp 589-599.

[60] Raji, A., Jeyasheeli, P. G. and Jenitha, T. (2016) "IoT based classification of vital signs data for chronic disease monitoring," *10th International Conference on Intelligent Systems and Control (ISCO)*, 2016.

[61] Munsell, B. C., Wee, C. Y., Keller, S. S., Weber, B., Elger, C., da Silva, L. A. T., Bonilha, L. (2015). Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. *NeuroImage,* 118, 219–230.

[62] Memarian, N., Kim, S., Dewar, S., EngelJr, J., & Staba, R. J. (2015). Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. *Computers in Biology and Medicine*, 64, 67–78.

[63] Dai, W., Brisimi, T. S., Adams, W. G., Mela, T., Saligrama, V., & Paschalidis, I. (2015). Prediction of hospitalization due to heart diseases by supervised learning methods. *International Journal of Medical Informatics,* 84(3), 189–197.

[64] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over Big data from healthcare communities. *IEEE Access*, 5, 8869–8879.

[65] Nilashi, M., binIbrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques. *Computers & Chemical Engineering*, 106, pp: 212–223.

[66] Kaya, Y., & Uyar, M. (2013). A hybrid decision support system based on rough set and

extreme learning machine for diagnosis of hepatitis disease. *Applied Soft Computing*, 13(8), 3429–3438.

[67] Khalid, Z., & Sezerman, O.U. (2018). Prediction of HIV drug resistance by combining sequence and structural properties. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* 15(3), 966–973.

[68] Raweh, A. A., Nassef, M., & Badr, A. (2018). A hybridized feature selection and extraction approach for enhancing cancer prediction based on DNA methylation. *IEEE Access*, 6, 15212–15223.

[69] Sedaghat, N., Fathy, M., Modarressi, M. H., & Shojaie, A. (2018). Combining supervised and unsupervised learning for improved miRNA target prediction. *IEEE/ACM Transaction on Computational Biology and Bioinformatics*, 15(5), 1594–1604.

[70] Wang, H., Cui, Z., Chen, Y., Avidan, M., Abdallah, A. B., & Kronzer, A. (2018). Predicting hospital readmission via cost-sensitive deep learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 15(6), 1968–1978.

[71] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, *7*, 81542–81554. https://doi.org/10.1109/ACCESS.2019.2923707

[72] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Computational Intelligence and Neuroscience*, *2021*. https://doi.org/10.1155/2021/8387680

[73] Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *Journal of Supercomputing*, *77*(5), 5198–5219. https://doi.org/10.1007/s11227-020-03481-x

[74] Stephen, M. (2014). *Machine Learning An Algorithmic Perspective Second Edition*. https://b-ok.cc/book/2543746/ef80cb

[75] Connor, C. W. (2019). Artificial Intelligence and Machine Learning in Anesthesiology. In *Anesthesiology* (Vol. 131, Issue 6). https://doi.org/10.1097/ALN.0000000000002694

[76] Panesar, A. (2021). Machine Learning and AI for Healthcare. In *Machine Learning and AI for Healthcare*. Apress. https://doi.org/10.1007/978-1-4842-6537-6

[77] Yadav, D. C., & Pal, S. (2020). Prediction of thyroid disease using decision tree ensemble method. *Human-Intelligent Systems Integration*, 1-7.