Cognitive Healthcare using Technology Integration

By

Muhammad Talha

00000327539

MS IT-2K20

Supervisor

Dr. Rafia Mumtaz

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Information Technology (MS IT)

In

School of Electrical Engineering & Computer Science,
National University of Sciences and Technology,
Islamabad, Pakistan.

(February 2023)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Cognitive Healthcare using Technology Integration" written by Muhammad Talha, (Registration No 00000327539), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: ______________________

Name of Advisor: Dr. Rafia Mumtaz

Date: 01-Feb-2023

HoD/Associate Dean: ______________________

Date: ______________________

Signature (Dean/Principal): ______________________

Date: ______________________

Approval

It is certified that the contents and form of the thesis entitled "Cognitive Healthcare using Technology Integration" submitted by Muhammad Talha have been found satisfactory for the requirement of the degree

Advisor : Dr. Rafia Mumtaz

Signature: ______________________

Date: 01-Feb-2023

Committee Member 1:Dr. Muhammad Ali Tahir

Signature:

31-Jan-2023

Committee Member 2:Dr. Tahira Lashari

Signature: ______________________

Date: 31-Jan-2023

Signature: ______________________

Date: ______________________

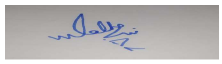
Dedication

I would love to dedicate my thesis to my beloved parents, teachers, my family, my friends and my supervisor Dr. Rafia Mumtaz who guided me throughout my degree.

Certificate of Originality

I hereby declare that this submission titled "Cognitive Healthcare using Technology Integration" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name:Muhammad Talha

Student Signature: 

Acknowledgments

In the name of Allah, the most beneficent, the most merciful. All praises to Allah Almighty for providing me courage to complete my thesis. I owe my deepest gratitude to my supervisor Dr. Rafia Mumtaz. Without her continuous concern about this work, her knowledgeable advice, support and encouragement this work would have been a far cry. I would also like to thanks my committee members Dr. Muhammad Ali Tahir and Dr. Omar Arif for their valuable advises during this research. I am incredibly thankful to my parents for their love and continuous support throughout this journey.

Muhammad Talha

Contents

List of Figures

List of Tables

Abstract

The Internet of Medical Things plays an important role in the healthcare domain for real-time monitoring of patients with high reliability and accuracy. According to the WHO in Pakistan, 30% to 40% of deaths are caused due to cardiac attacks which are approximately 200,000 deaths per year. A comprehensive literature study is conducted to explore, analyze and compare existing system architectures for cardiac health monitoring worldwide. Our preliminary survey shows that very few e-health architectures exist in Pakistan; therefore to address this issue, we proposed an digital health monitoring system that is able to detect the onset of various health anomalies in the patient's vitals, using advanced machine learning algorithms and data visualization using web portal. Thus, reducing the burden on hospitals by introducing remote monitoring facilities to patients as well as doctors. The fundamental purpose of the proposed research is to incorporate cutting-edge machine learning classification algorithms to detect anomoly in human vitals such as heart rate (HR), blood pressure (BP), blood oxygen saturation, body temperature, respiration rate etc. in near real-time. In our preliminary research, we evaluated the performance of multiple ML algorithms trained on the clinical data set. Random-Forest achieved the highest accuracy on the test set (95%) among the eight tested supervised classification algorithms. In order to provide a remote patient management and monitoring panel, we created an web portal to ensure the confidentiality and security of patient data within the proposed system.

Chapter 1

Introduction

Ever since technological advancements have been made, technologies like artificial intelligence (AI) have become much more powerful and are having a positive effect on human survival. AI helps all fields, including medicine, manufacturing, the military, and even everyday things like traffic lights, security cameras, voice recognition, and catching fraud, among others. Moreover, machine learning (ML) is essential since it enables robots to analyze and identify evolving behaviors. Efforts are underway to enhance the learning accuracy of ML models to create robots with human-level intelligence and learning capabilities. This involves utilizing deep learning, a subfield of machine learning that utilizes models based on natural neural networks to gain knowledge.

Using machine learning with both supervised and unsupervised models, artificial intelligence (AI) may be able to find abnormalities in human vital signs like heartbeat, temperature, blood pressure, oxygen saturation levels, and breathing rate. When patients are initially admitted to hospitals, their vital signs are recorded. This information is then maintained and kept up-to-date by the appropriate ward personnel. During the second phase, different machine learning algorithms were trained on the data to make an accurate prediction of cardiovascular diseases (CVD). The third step involved further training and optimization to increase the machine learning model's accuracy.

This field of study is relevant to the current state of the medical profession all over the World. As of right now, a significant number of deaths that have been reported are the result of cardiovascular disease (CVD), particularly in nations that are still developing. Issues relating to one's lifestyle, such as a lack of exercise or activity, a stressful job situation, financial troubles, a diet rich in cholesterol, and poor junk foods, are a big contributor to the rise in cardiovascular disease. According to a survey [50], approximately 200,000 people die in Pakistan each year as a result of heart-related diseases. People over 45 were thought to be more likely to die from cardiovascular disease (CVD) in the past, but recent deaths in younger people are raising serious concerns. Researchers have found that young people are more likely to die from heart attacks, which is a cause for concern. Being able to predict heart attacks can help all of humanity. CVD depends on human vitals as well. To survive, heart muscles required adequate oxygen levels in the blood. Blood is supplied to heart muscle by coronary arteries. On the other hand, the aorta is an artery that is used to carry blood from the heart to other body parts. The heart tissues will die quickly if the coronary arteries become blocked, as represented in figure 1.1. As a result, CVD diseases can develop.

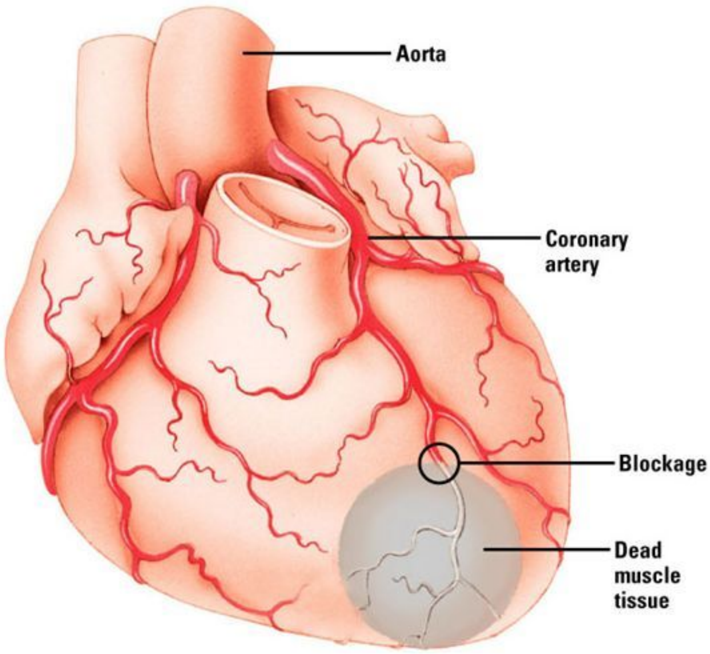


Figure 1.1: Working of Heart

Cardiovascular disease (CVD) is a chronic non-communicable disease (NCD) that, once it affects the heart organ or associated veins or capillaries, has effects that last a patient's entire life. These effects are immediate and have an effect on the patient's quality of life. One can avoid getting this condition by altering their lifestyle, and if they do, they can effectively treat it if it appears in the early stages to stop it from worsening. Before technology was invented, each patient had to be checked by hand. This meant that the doctor had to visit each patient and spend a lot of time getting their vitals and making a diagnosis. Modern technology like artificial intelligence (AI) and the Internet of Things (IOT), which can provide and analyze data in real time, has made it possible to create intelligent medical systems. These systems not only help gather vital signs in real time, but they can also use the information from the vital signs to diagnose conditions. These Internet of Things devices can track a patient's vital signs and see what he or she is doing. Depending on the patient, they can be attached to a patch, worn, or carried around. These gadgets might have storage, and they might send information to a hospital-based cloud server for additional processing. These data are analyzed and made accessible to healthcare providers so that they can make additional diagnoses and treatment suggestions [5].

Taking into consideration the IOT devices for collecting vitals and diagnosing, a patient can sometimes be less precise, which can lead to misleading results. As with any technology, there are positives and negatives associated with it. In any case, machines still can't fully replace human intelligence, but recent advances mean they can help and speed up the process of diagnosing. During the COVID-19 pandemic, numerous nations demonstrated the value of spending significant resources on artificial intelligence (AI) and internet of things (IoT) applications in the field of medicine. However, sadly, Pakistan is lagging behind in this field, and its absence is creating undue pressure on medical teams due to limited resources and an extreme number of patients. Pakistan's lack of progress in this area has caused Pakistan to fall behind other countries. We are able to reduce the strain placed on hospital medical facilities by utilizing this cutting-edge technology, which also enables us to improve patient monitoring and the tracking of vitals.

1.1 Problem Statement

The citizen of any nation has a fundamental right to access high-quality healthcare, yet many have fallen short in providing even the most essential treatments. The world and medical technology are both evolving. The rise in viral infections and chronic diseases such as heart disease, cancer, diabetes, and obesity has been accelerated by changes in lifestyle, societal issues, and environmental pollution. This is a significant problem for both Earth's life and our healthcare system. Hospitals, healthcare professionals, and pharmacy businesses all have to work harder to keep up with the demands of any new pandemic or endemic disease. Unfortunately, Pakistan doesn't have enough hospitals or medical supplies to deal with similar emergencies in the future, which could be very bad if they aren't handled correctly.

When it comes down to it, we can all agree that the World's population is growing at an alarming rate and that this has an impact on natural events. More people wish to reside in cities as they develop into metropolitan areas. Despite the fact that populations and cities are expanding swiftly, not nearly enough attention is paid to expanding medical facilities or modernizing those that already exist to accommodate the rising need. Due to a lack of resources and an increase in the number of patients, public hospitals are also unable to provide comprehensive, high-quality care to people with low incomes.

During the COVID-19 pandemic, the healthcare system was put under a lot of stress, which got a lot worse during the time when the virus was spreading the most. The increasing patient load was extremely difficult for people in the medical field, especially doctors and emergency medical personnel. To meet the needs, temporary hospitals with isolation rooms and intensive care units were built. These hospitals, though, also fell short of the actual demand. Additionally, it became increasingly difficult to assist those who had other types of health issues during COVID. In this instance, there was a new requirement for the remote provision of medical services, such as online examinations and diagnosis. This was a brand-new requirement. Several other systems emerged during this time, but none of them was able to offer a full solution to this issue. So, a system is needed that can track patients, diagnose them, and put them into groups so that only people with really serious or life-threatening illnesses go to the hospital. As a result, there should be less stress on the hospital's infrastructure.

1.2 Aims & Objective

Prior to presenting the problem statement, the job that must be completed will be broken down into its component elements. Collecting real-time data from hospitals regarding patients is the first step that must be taken. This information should include the patient's temperature, respiration rate, blood pressure, and heart rate. In the second step, the data are cleaned up by getting rid of useless information. This is done so that the data are ready for the preprocessing phase. In the third step, machine learning models will be trained and fine-tuned to get as close to perfect accuracy as possible. Also, the most accurate machine learning algorithms will be used to look for abnormalities in people's vital signs by analyzing their vital signs. For remote health monitoring, the web portal will be developed to view patient data at any time.

1.3 Organization of thesis

This chapter presented a general introduction and the objectives of the research work. Chapter 2 highlights the related work in this domain. Each research and review paper discussed in this chapter has been analyzed and studied in detail in order to find a direction for this research work. Moreover, this chapter also identifies the gap in the existing research that we have tried to overcome in this research work. The third chapter explains the procedure and methodology for this research. Additionally, the development of the prediction model involves the examination of supervised and unsupervised machine learning techniques, as well as the features of the data set, comparisons of various machine learning methods, and the difficulties encountered during deployment. The fourth chapter presents the findings result of this research project. In the fifth chapter, there is discussion about web portal for remote health monitoring and data visualization. The last chapter wraps up all of the research, shows where this work can go next, and lists its limitations.

Chapter 2

Literature Review

Recently, a lot of developments have been made in the healthcare field to provide state-of-the art solutions by employing cloud-based solutions, big data analysis, machine learning (ML), artificial intelligence (AI), and the Internet of Things (IoT) [44]. Due to COVID-19, a lot of research has been done and published on remote patient monitoring, which has been needed for the last ten years.

Access to quality healthcare is a fundamental human right that is rarely denied. Over the course of the previous century, there has been a considerable change in how people live, which has increased their susceptibility to a range of diseases. Experts in the field are also sure that monitoring and delivery systems for medical care made with the Internet of Things (IoT) can be helpful. There are three levels of IoT systems, with the first being the collection of patient vital sign data. The next step is to send all of this information to a service that stores data in the cloud. Thirdly, the Body Range Network aggregates the data provided by its sensors (BAN). The second layer is in charge of keeping records and using statistics to look at them when they reach a certain number. Also, algorithms for machine learning and data mining store data sent by the cloud and use it to do expensive calculations. The data collected by these algorithms can be used to determine what is wrong with a patient. The data that these algorithms collect can be used to figure out what is wrong with a patient. Should any issues arise, the data can be transmitted to the patient's physician through a web interface for closer examination.

Recent research, such as the study in [44], has focused on creating integrated systems for remote patient health monitoring. Every alternative that is theoretically possible is used to achieve this goal. The main objective of this research is to set performance standards and assess the available models and technologies for machine learning. The researchers also stressed the technology's potential real-world uses as well as how swiftly it could react. As part of this study [46], researchers have developed a system that involves ongoing patient monitoring while upholding a high level of respect for the patients' right to privacy. Deep Neural Networks (DNN) and Artificial Intelligence (AI) are used to find abnormalities in heart diseases. The system's main function is to store data on a cloud-hosted server. If a patient is getting worse and an emergency arises because of it, network delays caused by latency could kill people. This configuration directly slows down the system's overall reaction time by a significant amount. Patients' heart rates, blood pressures, and temperatures are all recorded in real time by biosensors at [25]. Biosensors with WiFi connectivity allow access to cloud storage. This solution does not offer analytical analysis of the collected data and is not adaptable enough to be used in the real world. As a result, it has very limited functionality and can only be used as a proof of concept.

Researchers can keep an eye on a person's heart rate with the help of a wearable smart device that has a photoplethysmography (PPG) sensor. In this illustration, a Raspberry Pi serving as the edge computing device analyses the data at first. The data is then moved to the cloud where it undergoes additional analysis. By utilizing a recurrent neural network (RNN) equipped with both long-term and short-term memory (LSTM), they were able to successfully forecast future heart rate. The individuals in charge of creating Telecare-ECG A monitoring system for cardiac patients developed by [41] can be used outside of a hospital setting. This device can be used to monitor an ECG through a wide range of gateway channels, such as Bluetooth, GPRS, and others. storing and analyzing ECG data in the cloud to detect heart irregularities The patient gets the EKG report through a mobile app, and a copy is sent to the doctor who is treating them.

The current system does not utilize machine learning for anomaly detection. Researchers at HealthFog are instead relying on deep learning (DL) and the Internet of Things (IoT) to create a healthcare system that can detect cardiac abnormalities [31]. They use a lightweight DL model that they term the "neural network" for the fog layer. In addition, they showed and analyzed the HealthFog implementation based on a number of performance factors, such as accuracy, reaction time, network bandwidth, and energy consumption.

The fundamental issue with the suggested system is that it lacks sufficient computing power to support DL model training and prediction. As a result, the creators of the app use traditional central processing units (CPUs) or graphics processing units (GPUs). This paradigm can't be used because edge computing devices can't do enough processing to make it work. Because of this, network latency affects how fast the system can respond.

The research project in [26] focuses on the design and development of a multi-parameter patient monitoring system that utilizes Internet of Things (IoT) technology. Using biosensors, vital signs of humans can be measured, including heart rate, respiration rate, oxygen saturation, and temperature of the body. In case of a critical event, an email notification is sent to the caregiver of the patient. The system's performance is improved by using Support Vector Machine (SVM), achieving 95% accuracy. Given that over 91% of the global population now owns a smartphone [7], it is believed by some that smartphone-based cardiac abnormality detection could provide a comprehensive, widespread, and cost-effective solution to healthcare. They used an accelerometer for monitoring the patient's activity, a magnetometer for monitoring the patient's movement, photoplethysmography (PPG) for monitoring the patient's heart rate, and a global positioning system (GPS) for tracking the patient's position. In order to identify irregularities in human vitals, the K-Nearest Neighbor (KNN) model was utilised.

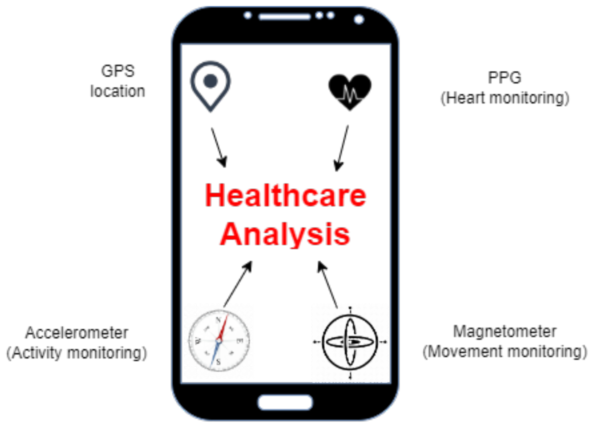


Figure 2.1: Proactive Healthcare Analysis [7]

The various methods are employed to find anomalies in human vital signs that are connected to cardiac health. These techniques can be categorized into the following sections, which are discussed in detail in this chapter:

- Machine & Deep Learning
- Cloud, Edge & fog computing
- Internet of Medical Things (IoMT)

2.1 Machine & Deep Learning

Researchers now have the ability to build and implement intelligent solutions because to the progression of technology. This section offers a summary of the systems that utilize machine learning and artificial intelligence to make informed assessments of a patient's medical condition.

The blood pressure and glucose measurements of patients are utilised by the [28] smart home health monitoring system, which then alerts medical professionals to any abnormalities in hypertension and diabetes levels. The evaluation of human patient measurement inputs in the system was performed using machine learning algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees (DT), Logistic Regression (LR), and Discriminant Analysis (DA) with conditional decision-making. The diabetes prediction algorithms developed by Sonar et al. are comprised of decision trees (DT), naive Bayes (NB), support vector machines (SVM), and artificial neural networks (ANN). The evaluation method used tenfold cross-validation, and values of typical prediction accuracy were kept. For its part, the LR method finished with the highest accuracy score (77.7%), while the coarse Gaussian SVM method had the lowest accuracy score (60%). For diabetes prediction, the study in [21] employed ML techniques including neural networks (NN), support vector machines (SVM), and random forests. The preprocessing techniques used included imputation, scaling, data normalization, and principal component analysis. The NN model performed the best among all models with 100 epochs and 10 batch sizes. It was accurate to a degree of 80.4% of the time. [19] used a deep neural network to predict diabetes and achieved an accuracy measurement of 77.86% overall. Scaling was used during the preprocessing stage. A ReLU activation function, a Nadam optimizer, one hundred iterations, and ten different batch sizes were used as configuration settings. During the review process, a technique known as "tenfold cross validation" was also used.

In this paper, [20], the researchers suggest a patient monitoring system that works with

IoT and machine learning. They argue that the essence of healthcare is to maintain health through the prevention and diagnosis of diseases. IoT is a potential solution to automate and optimize modern healthcare. Monitoring of patients is highly advantageous to make informed decisions. Their proposed framework includes a stage for patient supervision. It had a hardware section with sensors that could measure blood pressure, heart rate, and temperature and connect to a Raspberry Pi board. The cloud is used to store the sensor data, which is then examined to check for errors. Data analysis is performed using machine learning algorithms. Specialists can get the data and its analysis, and the patient's condition is posted on the hospital's website. The three main components of the proposed system are emergency alarms, health state prediction, and health monitoring. The module for predicting health state appears promising. A KNN classifier is used to sort the information stored in the database. KNN could be a non-parametric supervised learning technique in which the information is put into a certain category. A set of prepared data is used to train a model, which is then cross-validated to find the best K value. The designers were directed to achieve 98.02 percent accuracy on the test set. Sensor data was used for testing, and the UCI data set was first used for training and cross-validation. When it comes to health care and patient monitoring, the biggest issue is the confidentiality of patient data. Encryption mechanisms were used in this model but the The security of data in an IoT model remains a problem. Even the machine learning algorithm poses a limitation on the model. KNN can be executed quickly on a small data set, but it can cause a computational hindrance when the data set is large.

In this study of [9], a system based on the Internet of Things (IoT) and machine learning for tracking the well-being of elderly stroke survivors .The authors emphasise the need for vigilant monitoring of patients with chronic illnesses to guarantee that the correct drug is given at the right time. Almost all stroke victims die because they do not receive immediate, adequate medical care. Machine learning has been shown to be useful in the early stages of medical diagnosis. The current health of a patient is assessed through the use of appropriate sensors, with the resulting data transmitted to the cloud for storage. In case any of the parameters fall outside the normal range, an email and an immediate alert are sent to the responsible doctor or caregiver. The physician can access the patient's medical history or previous conditions through a web portal, and then update the prediction system with any new or pertinent information that they discover. The prediction model is made with the help of classifiers, and its goal is to find out if the patient has a high risk of having a stroke or not. The hardware layer of the suggested design is made up of a pressure sensor, a heart rate sensor, and a blood sugar sensor that all connect to an Arduino Mega microcontroller. Because the microcontroller incorporates Wi-Fi connectivity on its own, data may be sent directly to the cloud. In their writing, the authors were not able to make a strong case for either the microcontroller or the way they communicated. In addition to ensemble classifiers, more traditional classifiers like Naive Bayes, Decision Tree, K-Nearest Neighbor, and Random Forest were also researched during this time period. The outcomes demonstrate that the ensemble classifiers perform significantly better than their competition-level equivalents. The suggested model made use of random forest as an ensemble classifier, and it achieved an accuracy of 93% thanks to its implementation. This study did not make use of any data that was readily available to the public. The suggested model relies solely on data collected through the tracking of several key indicators. Cognitive impairment, weak limbs, trouble speaking, and going blind can all be signs of a stroke, but they are too vague to reliably predict a high risk of having one.

A healthcare monitoring system built on machine learning and controlled by an Arduino Uno and an Atmega328 microprocessor was designed by [26]. There will be two stages to complete the project. Even though the overall strategy of the classifier is discussed elsewhere, this chapter is mostly about making a prototype that works. In the first stage, an ESP8266 microcontroller uses a number of biosensors to measure and send information on four physiological parameters: heart rate, body temperature, electrocardiogram (ECG), and pulse oximeter. After that, they used the SVM machine learning method on the sensor data to make a prediction about the person's health. The SVM, or Super Vector Machine, is a fast and accurate option for processing data, with a system accuracy of up to 95%. In [30], a recursively improved version of the random forest (RFRF-ILM) is used to detect heart disease. Specifically, a method known as RFRF-ILM is used, which combines the benefits of both the linear model and the random forest. If the disease is caught early and a risk assessment for preventing disease is done as soon as possible, the number of deaths caused by the disease can be greatly reduced.

An "Enhanced Deep Learning Assisted Convolutional Neural Network (EDCNN)" is suggested in the [35]. The results of the tests show that the designed diagnostic system can accurately predict the risk of heart disease compared to other deep learning-based smart healthcare systems such as Artificial Neural Networks (ANN), Deep Neural Networks (DNN), Ensemble Deep Learning-based Smart Healthcare Systems (EDL-SHS), Recurrent Neural Networks (RNN), and the Neural Network Ensemble Method (NNE). This conclusion came from the analysis. EDCNN has a 99.1% success rate, which is really impressive.

Unsupervised machine learning anomaly detection was covered in depth in a tutorial made available by [42]. In the tutorial, participants get an understanding of the capability of unsupervised algorithms to classify aberrant behavioral events through participation in a hands-on session in which these algorithms are applied to a variety of data sets.

The study by [32] shows how machine learning can be used to produce personalised 10-year heart disease risk forecasts. The name of their classification algorithm is Hard Voting (HV). To build HV, standard classifiers such as Logistic Regression, Random Forest, Multilayer Perceptrons, and Gaussian Naive Bayes were employed. The meta-classifier was supplied with data on various potential predictors, such as age, sex, smoking status, systolic blood pressure, and others. The four classifiers are put to the test to determine which one makes the best prediction. For the investigation, a Kaggle data set with details on 3751 patients was employed. To normalise the data and get rid of anomalies, the scientists used RobustScaler. 80% of the total dataset was used for training, leaving 20% of it for testing. The creation of such a meta-classifier was successful, as shown by the evaluation results of 88.42% accuracy and 1.0 precision. The focus of this effort is on the development of a machine learning model rather than creating an IoT-based ECG monitoring system. Although a model has been created, it has not yet been evaluated using actual patient data. The suggested design not only collects a range of vital signs but also uses machine learning (ML) techniques to produce in-the-moment predictions about the health of patients. With the use of ML technology, diseases can now be identified and diagnosed early. This is essential in the case of chronic illnesses like heart disease. The results of this study support the notion that a customised machine learning approach could reliably forecast the prevalence of cardiovascular illness over the following ten years. "Hard Votes" are votes that are exceptionally difficult to pass (HV). HV was developed using the techniques of Gaussian Naive Bayes, Random Forest, Multilayer Perceptrons, and Logistic Regression. The meta-classifier takes into account systolic blood pressure, age, gender, smoking status, and other variables. To decide which of the four classifiers generates the most precise forecast, a vote is taken, and the results are added together. The scientists used a data collection with information on 3,751 patients that was made publically available on Kaggle. To remove any odd or unusual numbers from the data, the authors utilised RobustScaler to normalise it. An 80/20 split was utilised to divide the data set into two halves for training and testing. The fact that this meta-classifier technique passed the test with an accuracy of 88.42% and a precision of 1.0 shows that it was successful in its creation. Making an Internet of Things-based ECG monitoring system is not the major objective of this endeavour. Instead, creating a machine learning model is the primary objective. Using patient data gathered in real time, the authors have not implemented their model. The architecture under consideration uses machine learning (ML) approaches to produce real-time predictions about a patient's health.

CAMISA is an acronym for a plan to use artificial intelligence to monitor COVID-19 patients from a distance. The authors [43] developed a Wireless Sensor Network (WSN) system in the form of a smart shirt and nebulizer in order to collect information regarding the physiological features of patients. A patient's temperature, breathing rate, SpO2 level, pulse, and other vital signs are among the parameters that are measured by the gadget. In the event that any of these measurements go above their previously established limitations, notifications are generated. These notifications share the patient's present situation with the hospital in order to facilitate a speedy transfer to intensive care should the need arise. In addition to this, the authors have developed a neural network model that can predict the likelihood of a user being infected with a coronavirus. The Artificial Neural Network (ANN) model receives twenty parameters as input, one of which is whether or not the patient is experiencing a cold, fever, or cough—three separate symptoms that may be present during a coronavirus infection. The user enters these input elements into the health monitoring app with the assistance of a questionnaire. The software then decides whether or not the patient has COVID-19 based on the analysis result. Additionally, the app utilises the Thingspeak IoT platform in order to display the sensory data that is collected by the device. This tool can help hospitals in their attempts to stop the spread of the COVID-19 virus by enabling remote real-time monitoring of COVID-19 patients.

2.2 Cloud, Edge & Fog Computing

The most recent edge and fog computing concepts localise resources to offer end users low latency and energy-efficient options, bringing new improvements to cloud-based architecture. In this section, a list of the many methods that can be used is given.

The authors of this paper [37] have designed a health monitoring system using Arduino Mega. The system checks for three basic physiological parameters (Body temperature, ECG, and pulse rate) using standard body area network (BAN) sensors. The Data is then transferred to the Thingspeak server using the esp8266 Wi-Fi module. The The collected data is then displayed using the Arduino IDE and the Thingspeak server. The data representation is both numerical and graphical. However, a drawback of this study is that the prototype can only be used to display the data and not be of much use in emergency situations, as it has no way of dealing with emergency situations.

According to [25], there are two main parts to any IoT-based medical monitoring system: the sensor network and the communication network. Several physiological markers (blood pressure, ECG, SpO2, heart rate, pulse rate, blood fat, and blood glucose) and a continuous environmental indicator (the location of the patients) were planned to be taken at variable rates during the study. A prototype example was created and put into action to demonstrate how the system works. This monitoring system not only satisfied the prerequisites for ubiquitous healthcare for cardiac diseases, but it also accounted for the expenses associated with doing so, guaranteeing that the ubiquitous mode was implemented in the most economical way possible. An Arduino served as a central hub for all of the sensors used in the constant patient monitoring and data analysis. Ubidots web servers allow users to access and modify their projects remotely. Ubidots is a development platform that simplifies the process of collecting data from sensors and turning it into actionable insights for businesses. Any internet-connected gadget can use Ubidots to send information to the cloud. A DS18B20 sensor is housed inside of a protective enclosure so that its internal temperature may be read. Beginning at a cool 31.94 degrees Celsius, the temperature quickly climbs to a comfortable 36.50 degrees Celsius after just two minutes of measuring. In light of this, we can anticipate a considerable variation in temperature readings based on factors such as geographical location, season, and subject age.

Author propose a Fog (Edge Computing) based smart healthcare system called Health-

Fog for the autonomous diagnosis of heart diseases through the use of deep learning and the internet of things in the study of [38]. As a consequence of this, new research pathways can be investigated regarding the integration of complicated ensembles of deep learning models as a result of the fact that processing at the edge gives the primary benefit of reducing response time. They proposed a potential architecture for a computer system that may be utilised in the development of ensemble fog computing for deep learning. Their system, which they referred to as HealthFog, was able to analyse vast amounts of data from cardiac patients in a speedy and precise manner with to the assistance of ensemble deep learning. Real-time data processing was made possible as a result of their ability to interface HealthFog with the IoT Edge-cloud, which was made possible by the architecture of FogBus. For the HealthFog configuration, a range of performance parameters including accuracy, reaction time, network bandwidth, and energy usage were displayed and analysed. In order to determine whether or not a patient has a heart problem, every piece of patient data that was accessible was analyzed. Using computing at the edge, which enables the production of results with a high degree of precision in real time.

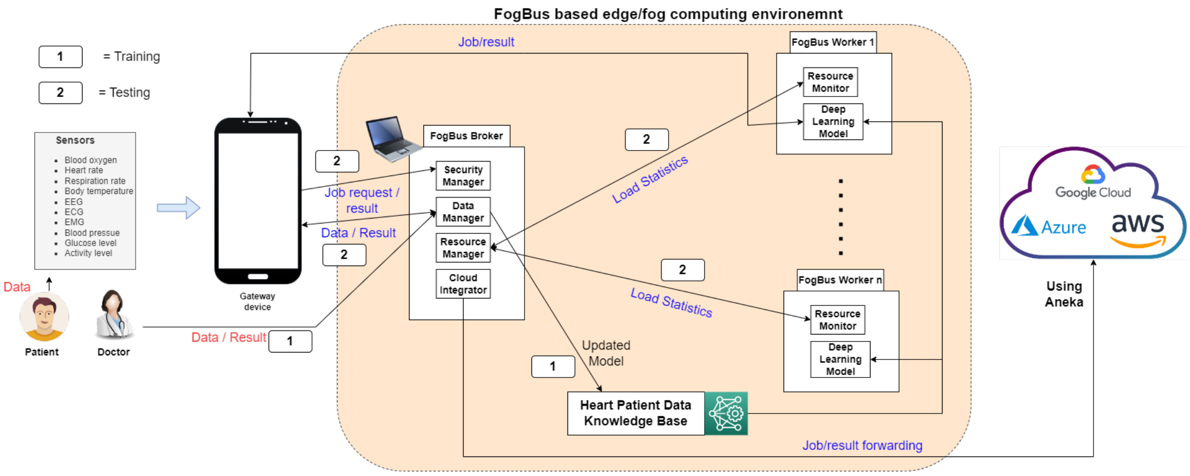


Figure 2.2: Fog Bus [38]

Machine learning has the potential to be applied in the diagnosis and prognosis of disease, as stated in the study paper referenced at [39]. They employ classification techniques like the random forest algorithm and K-nearest neighbours (KNN) for diagnosis and forecasting. All of the sensor data is collected by the cloud system, which means that it takes longer to process the data and get the result that was wanted. Computing in the fog and at the edge, which are located close to the IoT device and require the least amount of time to store and compute data, is favored as a solution for getting over this problem. In addition to this, they cut down on the amount of time that is needed to transfer data from an Internet of Things device to a storage system. Fog computing can help the health care industry run more smoothly.

The aim is to enhance an Internet of Things monitoring system [29] by incorporating real-time heart rate monitoring and analysis, and comparing the efficacy of PPG sensors in smart wearables to other well-established heart rate sensors in a clinical setting. After that, the measurements are sent to the application so they can be processed ahead of time. Send the measurements, which have already been preprocessed, to the cloud so that they can be analyzed there. A abrupt increase in heart rate is not picked up by the Fitbit sensor when the wearer is asleep. After that, the data is put to use in order to train the RNN model, which is then used to make predictions regarding HRV levels and B2B simulation.

An additional intelligent data analytics architecture for e-healthcare applications is presented in the article [33]. Federated learning is used in this design using regional edge computing hardware. Federated learning is a crucial component of the system architecture because it protects the privacy and security of private medical information gathered by wearable technology. This data can only be obtained successfully if federated learning is implemented. The study gives a thorough analysis of the suggested framework, which can be broken down into three parts: the cloud module, the edge module, and the application module. But the researcher don't give a detailed explanation of how the system works in practice. The major objective of this study was to determine whether federated learning might be used as an edge case in the design of e-health systems to guarantee service quality (quick reaction times) and data privacy.

2.3 Internet of Medical Things (IoMT)

The Internet of Things is now very significant in the healthcare industry. The term "Internet of Medical Things" refers to how medical software and hardware can work together over the internet. IoMT's work flow is depicted in reference figure IoMT. It might have sensors built into the human body. A smart watch is also used to measure health. In addition, smart phones include a variety of apps for tracking health. In the Internet of Medical Things (IoMT), there are various domains. The following are some domain areas:

- Tracking of medical equipment
- Smart hospital environment
- Remote monitoring of medical parameters
- Look after of old people
- Disabled people Assistance
- Personal monitoring
- Tracker for new born babies

A patient's body can be hooked up to a temperature sensor, an ECG sensor, and other medical sensors. Furthermore, smart appliances such as smartphones and smart watches contain embedded sensors. Edge computing can also be performed for immediate response. Based on the data that has been collected, if the data is important, it is sent to the server with an alarm, and if it is not, it is sent to storage. Gateway can be used to send data to the cloud for storage.

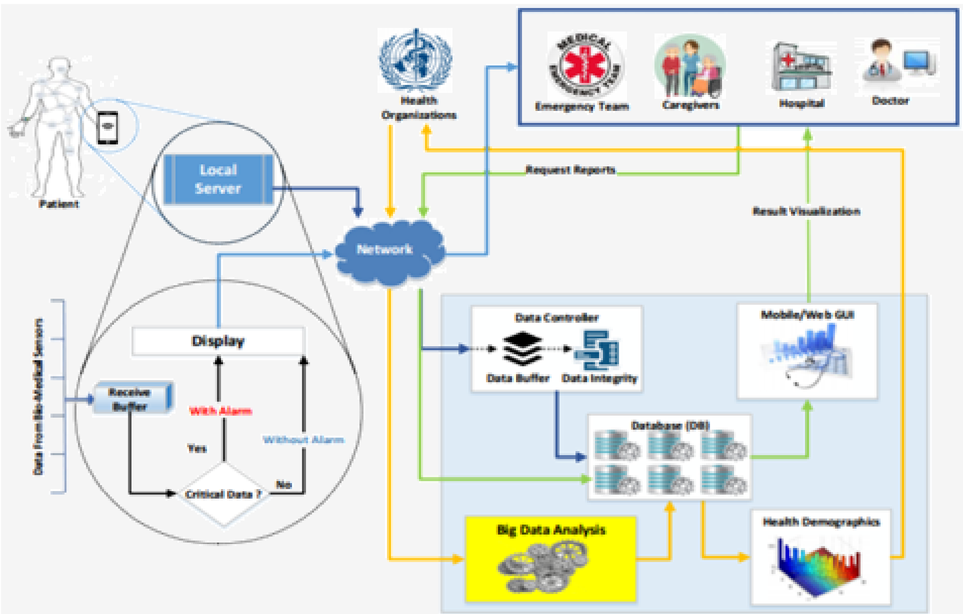


Figure 2.3: Work flow of Internet of Medical Things

Information about the patient can be sent to a wide range of medical professionals, such as the team in the emergency room, the patient's family, hospital staff, and doctors, for analysis. Data can be shown in a lot of different ways in mobile and web apps. Graphs and tables are two examples. In the not too distant future, artificial intelligence (AI) and machine learning algorithms, when applied to patient data, will make it possible to predict patients' health.

Other countries than those in South Asia have made more progress in the creation of human health monitoring systems that are based on the Internet of Things. Long-distance transmission and remote medical consultation have come a long way. Rezaeibagha and Mu constructed an agent system with wireless sensors that can monitor vital signs such as blood pressure, pulse, respiration rate, and body temperature [16]. A remote video medical diagnosis system has been developed by Agnisarman et al. [8]. This system makes use of the Internet of Things in order to make the process of medical diagnosis by a doctor easier. It can be used in different ways, such as through online consultation and video chat. In order to facilitate remote monitoring, Mugica et al. [12] developed an electrocardiogram (ECG) monitor that is compatible with the intelligent terminal of the Android system. Using the IoT platform, Tamilselvi et al. have come up with a plan for a health monitoring system that can look at important signs like a patient's oxygen level, body temperature, and eye movement.

In order to streamline the process of a doctor giving a medical diagnosis, researchers at [8] have created a remote video medical diagnostic system. Multipurpose platforms like online consultation and video chat are used to achieve this. Using the IoT platform, Tamilselvi et al. came up with a plan for a health monitoring system that can check important signs like a patient's oxygen level, body temperature, and eye movement. Acharya and others, as well as Tamilselvi et al., recommended this system. The data visualization tool is meaningless because the system does not have an interface. The system has a significant issue, and it is this issue. Banerjee et al. [6] created a technique that measures heart rate without the use of any invasive treatments. This solution uses a platform that lets interactive Internet of Things applications be watched in real time. An interactive smartphone application was created by Gregoski et al., Oresko et al., [4] to track the user's heart rate over time and identify cardiovascular diseases. Trivedi and colleagues came up with a mobile way of monitoring analog data for the purpose of conducting surveillance. Sending physical amounts to the device requires Bluetooth connectivity, and the Arduino platform handles any digital conversions that are needed. Kumar et al. [11] proposed three distinct IoT platform layers for their safety gadget:

control, device, and transport. Both Wi-Fi and Ethernet are used to send the data to the cloud platform. A wireless sensor network-based method for tracking smart homes and cardiac monitoring was presented by Desai et al. [10]. In their investigation, they used the Spartan3 and FPGA interface.

In order to track the subject's heart rate and body fat percentage, the system architecture used in the paper [22] makes use of an ARM-based Internet of Things development board that is interfaced with a number of sensors. The suggested system uses a temperature sensor to gauge body temperature, an electrocardiogram (ECG) sensor to gauge heart rate, a body fat analyzer to gauge percentage of body fat, and an LCD display to show information to the user. The development board has a Wi-Fi module that lets you send data to a real-time database. The database can be accessed by both the patient and the healthcare professional, but only after they have proven who they are. Therefore, the data can be used by the doctors to assess and monitor the patient's health.

In their work [18], which was released in 2018, the authors offer a WISE system for IoT-enabled health monitoring. The Body Area Sensor Network's architecture now includes a variety of wearable sensors (BASN). Three different types of sensing apparatus are used by the WISE body area network: sensors that keep track of blood pressure, body temperature, and the rate at which the heart beats. The authors say that most wearable healthcare solutions on the market today use smartphones to process, display, and send the data they collect. After the advent of these technologies, which demand ongoing supervision, how people use mobile phones in daily life will be very different. The authors suggest using WI-FI to send sensor data to the cloud so that it can be saved and then processed later. There isn't a strong case for why the writers chose the type of communication technology they did. On a small, lightweight LCD screen, the user may see their current data, including their temperature and heart rate. An RFID reader is used to verify the identity of the user. Patients, medical professionals, and caregivers can all look at the data trend and its history on a website. A learning technique like support vector machines (SVM) or neural networks is employed in the decision-making process after applying data mining to extract meaningful qualities from the data. Since the authors didn't say what algorithm they used or how accurate it was, this seems like a big thing that was left out of the report. Nevertheless, any anomalous readings are immediately reported.

This prototype's compute and storage capacities are severely constrained due to its reliance on a single method of connecting to the cloud. This is a major drawback. Since there isn't a device at the edge to process data before it's sent to the cloud, the massive amounts of information collected by sensors aren't cleaned up properly. This can add complexity to the data and lead to a high number of incorrect measurements. But because this journal doesn't give any numbers, it is impossible to judge how important the research described in it is. That's the main issue with the way it was built. More work is also needed on the data mining component's development. The machine learning algorithms utilized by this system should be thoroughly analysed.

The microcontrollers in the health monitoring system discussed here are ARM Cortex models. ECG, blood pressure, pulse rate, body temperature, body posture, and a live camera are the seven basic ways to check a patient's health. An Ethernet cable is used to transmit the information to the server. For security reasons, we have assigned a different IP address to each user; this address is what unlocks the door when they log in. The IDE used to create this system is L. Attolic TrueStudio (IDE). However, this technology is limited in its scope due to its inability to send or receive wireless data.

[40], an Arduino-based healthcare monitoring system has been developed to aggregate data from three different sensor kinds (a body temperature sensor, a heart rate sensor, and a room temperature and humidity sensor). Analog information is digitalized with the help of a microcontroller. The information is then sent to the server using the standard chosen by the framework (in this case, ZigBee). The information is then wirelessly transferred to a local therapeutic server. The person's health is figured out by comparing all of the important information to norms that have already been set. The readings from the sensors are shown on both the LCD screen of the prototype and the Blynk app. This concept has flaws since it does not incorporate enough sensors to monitor a person's vitals and does not use artificial intelligence to make choices. This is very important, especially for people who live in remote areas where there aren't many medical facilities.

In the publication [23] researchers presented a state-based synchronization mechanism termed "echoes." The main goal is to provide a safe way to communicate in case of a connection break, and the conflict management mechanism tries to make sure that data is reliable when the connection is back up. The 'ECHOES' system incorporated a secure data connection and a privacy management platform.

The author thinks that by 2020, 50 billion smart devices will be able to support several services. Every intelligent device and service provider has its own data processing and storage methods. IoT can only reach its full potential when all of a person's data is available on a smart device, which can't happen without a unique protocol. To do this, the Echo protocol for secure data management and communication is being made. The procedure has the following key characteristics:

- **Synchronized central server:** If any data changes, the server will apply new changes to all smart devices.
- **Offline mode:** If there is no Internet connection, data should be stored locally in a database. After a successful reconnection, the data should be synced.
- **SDC:** Secure data connection
- **PMP:** Privacy management platform in mobile app

The author focused on smart health and a healthcare use case in the study. In this case, biosensors find out how healthy the patient is and tell the doctor about it. Medical experts can now use cellphones to make diagnoses, so patients no longer have to go to the doctor or the hospital unless absolutely necessary. In the same way, a case study showed how GPS was used to keep an eye on a diabetic patient.

The Internet of Things is based on distributed computing, which needs a data exchange system that is scalable, reliable, and always available. IoT data management requires the following components:

1. **IoT Data Availability:** Data should be available on all smart devices.
2. **Conflict Handling:** on synchronization of data if conflicts happened on the same data by editing the data by multiple devices in offline mode, the conflict must be resolved immediately.
3. **Data efficiency:** usage of resources should be limited at high speed.
4. **Transparency:** During the data synchronization process, there should be no user interaction, and the process must be invisible to user.

5. **Genericity:** The mechanism is not specific to any data or domain.

6. **Scalability:** Data must be synchronized across all smart devices and applications of IoT.

7. **Security:** Data privacy must not be jeopardized in any aspect of data processing, including storage, access, and sharing.

According to requirements that mentioned above, three main approaches are considered for android based smart things:

1. **Smart Devices:** First time, Android support facility as content provider to facilitate interfaces to every application but in result genericity loosed. For making genericity, MetaService established for temporarily storing of data by Secure Data Container in smart phones. But it did not support large data. At last, NoSQL database provided by CURATOR for sharing IOT data to applications directly and no need to store in smart devices.

2. **Cloud based:** In cloud based approach, cloud provider support facility to create database at cloud and smart devices directly access cloud database for storage of data. In cloud approach, Internet must be required and data can be loss due to disconnection.

3. **Data Synchronization:** In this approach middleware is used for data synchronization. It can handle conflicts in data but not effectively, and it is highly expensive.

In light of the aforementioned requirements and scenarios, ECHOES provides mechanisms to meet the needs of smart things that can operate in both offline and online modes and handle data conflict when reconnected to the internet. They make two synchronization processes: one that works in offline mode, and another that works in online mode.

For testing their proposed ECHOES mechanism, they used two smart phones and a Raspberry Pi as a server, as well as SQLite databases on the smart devices and MariaDB on the server. They evaluate amazing result that in 0.4 seconds changes in IOT data available to all clients.

The paper [27] discusses the significance of an IoT architecture in a densely populated third-world country like India. The healthcare services are not up to the mark, especially for those living in rural and remote areas, and there are not enough doctors in the country to handle the ever-growing population. The authors talk about how important it is for IoT to help make healthcare services better. This architecture may incorporate internet-based services, mobile services, text messages, and call centre services. The proposed architecture includes: sensors and actuators like wearable devices (smart watches or bands) or even devices like pulse observation sleeves, glucometers, oximeters, etc. that can be used for remote monitoring. Gateways are needed for Data acquisition to transmit the huge amount of data from these sensors. Edge IT is required to preprocess obtained data prior to its transmission to the cloud. The data center or cloud is where this data is stored and where analysis can be done on it. Management services and security are provided for the data here, and the data is then used to propose or suggest to the patients how to proceed. They are also promoting the use of electronic health records, which can assist IoMT in tracking and providing better insight to the patients. It will be important to develop a variety of monitoring-oriented applications and software that are easy to use. Measurements of fundamental bodily functions, such as core blood sugar, body temperature, body weight, and blood pressure. Hence, IoMT can be used for remote monitoring, medical alerts, improving medical tests, removing human errors, fast and better identification of diseases, and overall better healthcare services.

An on-off algorithm tailored to medical data is discussed in the article [15] which also emphasizes the importance of the IoMT and places special emphasis on the energy efficiency of these devices and the system. By following the duty cycle theory, the transceiver radio is turned off when it is not in use. Overall, this study's findings are significant because they point the way toward an effective algorithm for transmitting data while minimising energy use.

This research also shows an innovative three-layer framework for the Internet of Medical Things. The first layer is made up of the various medical equipment that is employed in the process of data collection. These can include things like wristbands, fitness bands, thermometers, pulse meters, glucometers, and heart-rate monitors. The integration layer is the second layer, and it's the one in which the devices are integrated with particular protocols of the Internet of Medical Things and transmit the data using the energy-efficient On-Off algorithm. The third layer consists of not only storage and remote

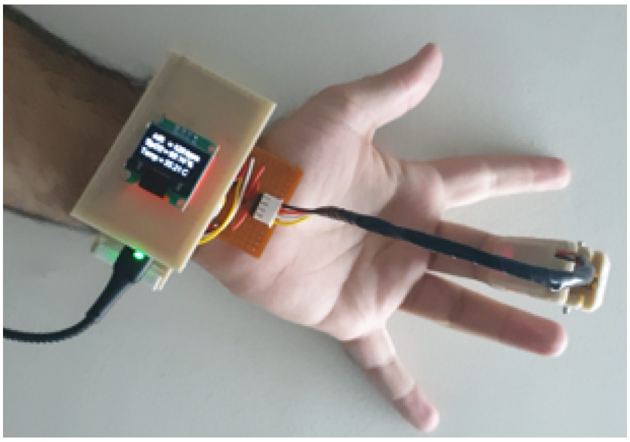


Figure 2.4: IoMT Node [47]

monitoring tools but also acts as a guide and support for doctors in the future.

A patient monitoring (PMM) system based on the Internet of Things (IoT) was designed by [26]. This system employs sensors to continuously monitor four vital parameters (heart rate, respiration rate, oxygen level, and temperature). In case of an emergency, an email alert is sent to the caregiver assigned to the patient. As part of this effort, the Support Vector Machine (SVM) algorithm will be utilized to enhance the MPM system's performance. A classification accuracy level of 95% has been determined.

[47] proposes an AI and edge computing-based IoMT-based remote patient monitoring system. This technology would allow for remote monitoring of patients. Developing an embedded prototype for remote monitoring of cardiovascular patients is the primary goal of this research, as shown in figure 2.4.

Users who have been verified will get real-time updates on their health based on the system's constant monitoring of vital signs like core body temperature, heart rate, and oxygen levels in the blood. Edge computing is used for many things, such as sending data between a sensor network and the cloud and sending out alerts in case of an emergency. The system also leverages a machine learning (ML) model to anticipate the user's health state, based on current data. Access to the heart rate, temperature, and oxygen level information is provided in real-time. A web application has been created to present the raw data and the outputs of the machine learning model. This will make it easier for the patient and doctor to communicate. When using the K-Nearest Neighbors (KNNs) technique, the machine learning (ML) module scored 96.26 percent correct on the test set. The plan's purpose is to make people feel less harried. The shocking news about the death rate of people with cardiovascular disease is making people feel more and more like they need to act quickly. The project will pave the road for an intelligent IoT and ML-based solution that will improve people's quality of life and, ultimately, save their lives.

The authors [45] developed and evaluated a variety of quality-of-life applications centered on the home environment that make use of technology from the Internet of Medical Things (IoMT). In the creation of deep learning applications that will provide a user-friendly COVID-19 management platform for home use, a variety of edge-compatible hardware was utilized. These programs have been updated so that they can be used with edge-based inferences. They can also be used on mobile devices and web-based clients. The computer architecture gathers data from a variety of streams, such as data from sensors, feeds from cameras and videos, and thermal pictures. Using their own code, the authors show how a group of GPUs can be used to train a number of deep learning applications. After that, these applications are put through several tests to determine whether or not they are feasible. The following applications were tested and evaluated: cough sound analysis from cellphones, drowsiness detection, face mask detection, ECG signal classification, emotional analysis for determining physiological states, fever detection using thermal imaging, and emotion recognition using an EEG signal. In the publication, it is indicated how accurate the training and validation for each of these are, and the data indicate that they are accurate. Also, their system uses alert generations to help manage and give health care at home, especially to the elderly.

The importance of cardiovascular disease monitoring systems in general is brought to light by [24], which stresses the necessity for medical practitioners to have access to data from patients living in rural areas. People who live in rural areas are usually far away from people who live in cities, so they don't always have access to modern medical facilities. This paper explores the design and implementation of a wearable Electrocardiogram (ECG) monitor that works with Arduino. The architecture of the system involves bio-sensing modules, a cloud platform (Google Firebase) for real-time patient data storage, and a web interface that enables physicians to swiftly detect and treat any developing cardiac issues. The concept and implementation of the system are thoroughly discussed in the paper. This project will produce a low-cost, portable Internet-of-Things-based system for acquiring, transmitting, storing, and displaying electrocardiogram (ECG) data in real time. This system will be all-inclusive and work in real time. Professionals in the medical field can look at this kind of data in real time to figure out if someone has a heart disease. Also, the information is sent to a diagnostic program that uses a graphical user interface (GUI). The biggest problem with using this technology is the trouble with network access and latency, especially in more remote places. To solve this very important problem, they came up with a solution with an edge-based architecture that uses machine learning to find differences in the local area. This makes it easier to respond to urgent situations.

The authors of the research article [36] suggest a solution to the problem of remote monitoring in the form of a system that is based on the Internet of Things and machine learning technology. To do this, it builds a sensor network by adding Arduino to the system and using parts from the Internet of Things (in this case, a pulse sensor). This system makes use of data classification methods based on machine learning. These methods are then used to classify cardiovascular diseases in greater detail using the data that has been gathered. Based on the recorded data, this system also employs machine learning classification algorithms to classify heart diseases. The decision tree method, the random-backwoods classifier, and the support vector machine techniques are all used in this study (SVM). The above-mentioned algorithms were made by the authors using data samples from forty different patients. It was found that SVM is better at figuring out if there is a problem with the heart. Also, this was taken into account. If the proposed hardware and software solution is put into place, patients will have a better idea of when to expect cardiac symptoms to start. Although this approach can help patients with heart disease, a diagnosis based solely on pulse readings does not provide a complete picture of the patient's condition. Including a wider variety of sensors in the ML model's setup can boost its overall effectiveness. This is why they implemented a suite of sensors to monitor the chronic patient's vitals in the system they designed.

In this study, a system is proposed that can find problems with a person's heart health by looking at their vital signs. Additionally, patient data is kept in the cloud for future access and evaluation. It is additionally employed for data visualization on web portals. After a thorough examination of available options, the machine learning model best suited for identifying anomalies in human vital signs was chosen.

2.4 Comparative Table

The comparison summary of literature is shown in table 2.1. It represented the human vitals and the advanced technologies they used in their research.

Citation	Vitals					Edge Computing	Cloud Computing	ML / DL	Web / Mobile App
	Heart Rate	Body Temp	BP	SPO2	ECG				
[46]	✓	✓	✓	✓	✓	✓	✓	✓	✓
[41]					✓	✓	✓		✓
[38]		✓		✓	✓		✓	✓	✓
[29]	✓				BCG	✓		✓	✓
[25]	✓		✓	✓	✓	✓	✓		✓
[26]	✓	✓		✓	✓	✓		✓	
[14]	✓	✓		✓	✓		✓		✓
[47]	✓	✓		✓		✓	✓	✓	✓
[18]	✓	✓	✓				✓	✓	✓
[13]	✓	✓				✓			✓
[17]	✓	✓				✓			✓
[22]	✓	✓				✓			✓
[20]	✓	✓	✓			✓	✓	✓	✓
[9]	✓		✓			✓	✓	✓	✓
[37]	✓	✓			✓		✓		✓
[34]	✓	✓							✓

Table 2.1: Literature review comparative table

Chapter 3

Methodology

A machine learning-based system is proposed in this study to discover anomalies in human vitals that are associated with cardiac health. The architecture proposed in the study is shown in Figure 3.1 and consists of three phases: data acquisition, preprocessing, and machine learning models. These phases will be further discussed in the following section.

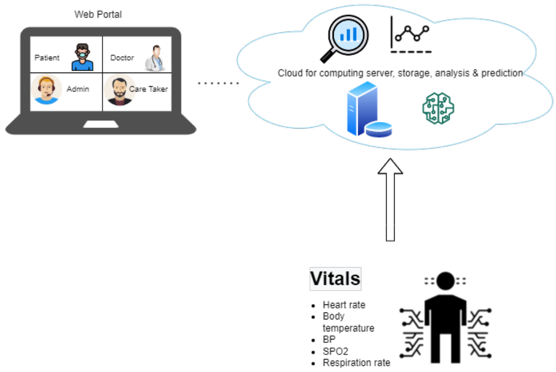


Figure 3.1: Architecture Diagram

3.1 Data Acquisition

The dataset that was utilised for the purpose of this research was obtained from the MIMICIII clinical database [48], which is housed on Amazon Simple Storage Service (S3). The MIMIC-III database is a big source of medical information that anyone can use for free. It contains information on 40.000 patients who were being treated in intensive care units at various hospitals throughout the country. The database holds important medical data related to cardiac wellness, including blood pressure, heart rate (in beats per minute), body temperature (in degrees Celsius), respiration rate (in breaths per minute), blood oxygen saturation (spO2), and other relevant information (in mmHg). The subject's cardiovascular health is going to be monitored, so the dataset that was retrieved contains all of the pertinent medical information that is required for this. In particular, the number of heartbeats that occur in one minute (BPM), the temperature of the body in degrees Celsius, the number of breaths that occur in one minute (BPM), the blood oxygen saturation, and the blood pressure (mmHg).

3.2 Data Pre-Processing

The extracted data-set will be preprocessed and calibrated by the system before ML can be integrated into the proposed system. This is a prerequisite for the integration. Utilizing processes that are considered standard for data preparation This is done to ensure that there are no training errors in the models used for machine learning. During our investigation, we partitioned the dataset into two distinct groups: healthy and unhealthy or '0' and '1', as seen in the figure 3.2. The collection of data was then labelled in accordance with the typical range of human vitals, as specified by [49]. The ranges that were used are listed below; they are as follows:

1. **Heart Rate**: 60-100 BPM
2. **Body Temperature**: 97-99 F
3. **Blood Oxygen Saturation**: >95%
4. **Respiration Rate**: 12-16 BPM
5. **Systolic BP (SBP)**: <120 mmHg
6. **Diastolic BP (DBP)**: <80 mmHg

It is essential to note that, in order to simulate the physiological activity of the human body, a 10% mistake was inserted while data set labeling. The data division ratio for training and validation was maintained at a predetermined ratio of 75% to 25%.

	heart_rate	body_temp	spo2	resp_rate	sbp	dbp	Label
0	74	98.1	100.0	24.0	149.0	65.0	1.0
1	71	97.7	98.0	12.0	101.0	70.0	0.0
2	72	98.6	100.0	19.0	110.0	70.0	1.0
3	71	96.7	98.0	24.0	144.0	51.0	1.0
4	60	97.3	95.0	15.0	133.0	46.0	1.0

Figure 3.2: Labelled Data set

3.3 Supervised Machine Learning Models

In supervised machine learning, a model is trained using labeled data, which consists of inputs and corresponding outputs. The goal of the model is to accurately predict the output for new or unseen inputs based on the patterns learned from the labeled training data. The prediction is done using the pattern the model learn during training. The data format must be matched with training dataset.

To train a supervised machine learning model, you need a dataset of input examples and corresponding correct outputs. You can then use an algorithm to adjust the model's parameters so that it can make accurate predictions on the training data. The algorithm works by comparing the model's predictions to the actual outputs and adjusting the model's parameters to minimize the difference between the two. This process, known as training, continues until the model's predictions match the actual outputs as closely as possible. After training, the model can then be used to make predictions on new data.

According to the findings of our study of the relevant literature, a number of different classifications of ML algorithms are used to identify and to ascertain the disease and diagnosis of cardiovascular diseases. Because of this, we chose the best models to use in our pilot study.

The workflow of supervised machine learning models is represented in figure 3.4. The dataset contains different human vitals like heart rate, body temperature, blood pressure including both systolic and diastolic, blood oxygen saturation level, and respiration rate. After data acquisition, data pre processig is required. For example, the removal of unusual data from raw data and data labeling. Before training of machine learning models, the dataset is divided into two subsets. The first dataset is utilized to train various machine learning models, while the second dataset is used to validate the performance of these models. To assess the models' performance, various evaluation metrics are used, such as accuracy, precision, recall, f1 score, and confusion matrix. The trained models are used to predict the health status of patients as either healthy or unhealthy.

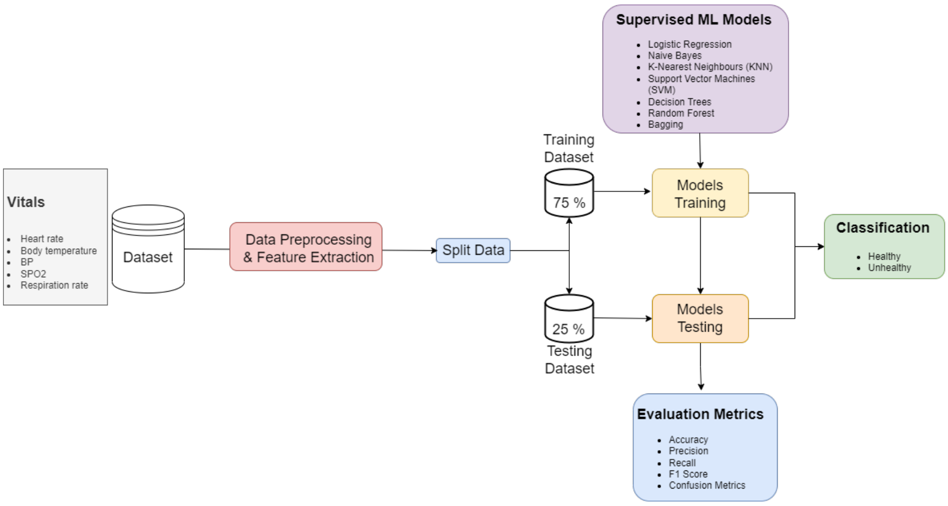


Figure 3.3: Supervised Machine Learning Models Workflow

The following is a list of several algorithms that are applied frequently in machine learning for cardiac health monitoring:

1. Logistic Regression

   The Logistic Regression (LR) is a statistical analysis method that predicts the probability of a binary output variable occurring, given one or more input variables [**mccullagh2019Generalized**]. This is an extension of the linear regression model, which include a dependent and a independent variable. The equation 3.3.1 explains how the algorithm evaluates the link between the variables used in the input and those used in the output by applying the model that provides the best fit to the data.

$$P(y \mid x, w) = \frac{1}{1 + \exp\left(-y\left(w^T x + b\right)\right)} \tag{3.3.1}$$

   The probability model makes a prediction of the presence or absence of a disease based on the input data and weights. The optimal weights are calculated by minimizing the negative log-likelihood function, as stated in equation 3.3.2, for each input instance in the set of training instances. The equation 3.3.2 is used to determine the optimal weights for the ith input instance.

$$\min_{w,b} \sum_{i=1}^{N} \log\left(1 + \exp\left(-y_i\left(w^T x_i + b\right)\right)\right) \tag{3.3.2}$$

2. Naive Bayes

   The Naive Bayes prediction technique assumes a strong independence between the feature qualities of the data. It is based on the Bayes theorem of probability and uses the input feature attributes for cases in the training set to determine the likelihood of an output prediction category. The Bayes theorem, as stated by Equation 3.3.4, is used to calculate the likelihood of event A given that event B has already occurred. The likelihoods of event B given that event A has already happened and the likelihood of both events happening are also considered.

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{3.3.3}$$

   The feature values in this work are continuous, therefore a Gaussian distribution is utilised to represent them in equation 3.3.4 This is demonstrated by the equation, which reads as follows: where represents the probability, where represents the parameterize mean and variance of all quantitative features, and where represents the ith attribute and value of all data instances. As the equation demonstrates equation 3.3.4, the classification of the anomaly in the vital signs has a role in determining the consequences of its presence.

$$p\left(X_i = x_i \mid C = c_j\right) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\left(X_i - \mu_{ij}\right)^2}{2\sigma_{ij}^2}\right) \tag{3.3.4}$$

3. K-Nearest Neighbours (KNN)

   KNN, or k-nearest neighbor, is a supervised machine learning method that classifies instances by analyzing the majority vote of their K nearest neighbors in the saved dataset. This algorithm is simple and has a fast convergence rate, and the similarity between instances is determined by a chosen distance metric, with Euclidean distance being the most commonly used. The number of neighbors (K) used in the analysis can be adjusted, with the value of 20 being found to provide optimal performance in this study. The Euclidean distance between two instances is represented by Equation 3.3.5 with N representing the total number of data instances.

$$d\left(x_i, x_j\right) = \sqrt{\sum_{i=1}^{N} \left(x_i - x_j\right)^2} \tag{3.3.5}$$

4. Support Vector Machines (SVM)

The Support Vector Machine (SVM) [2] is a supervised learning model that performs predictions by dividing data instances optimally and linearly using support vectors. For binary predictions, the algorithm forms two hyperplanes with maximum margins, which separates the training instances of the defined classes with the highest degree of accuracy. These hyperplanes are based on the number of feature attributes. The SVM maps each data instance as a point in a high-dimensional space, with the number of dimensions equal to the number of feature attributes. The data instances that are located in the closest proximity to the hyperplanes are referred to as support vectors, and new instances are placed in a category according to the relative distance that separates them from the hyperplanes. In this work, the radial basis kernel is used to the SVM in order to produce the SVM-RBF, which is then used to predict outcomes of non-linear data using the SVM by translating input features into a higher-dimensional feature space. The optimal hyperplane margins and their related constraints are located by the equation 3.3.6, which defines the minimization function utilized. A feature vector that corresponds to each data occurrence, a normal, a threshold, and a value that defines the threshold.

$$\min_{w,b} \frac{1}{2}\|\vec{w}\|^2 \text{ subject to the constraints } y_i\left(\vec{w} \cdot \vec{x}_i + b\right) \geq 1 \tag{3.3.6}$$

5. Decision Trees

The decision tree (DT) [1] is one of the most well-known and widely used machine learning methods for classification and regression. It is a supervised machine learning approach in which building a training model for predicting the class of target variables takes into account decision rules rather than training data. The method begins at the tree's root, and depending on the direction we move away from it, we compare the values of the root property with those of the record's attribute. The sample's complete population is represented by the root node. Two or more sub-nodes can be created from a single node. A sub-node is regarded as a decision node if it divides into other sub-nodes. Leaf nodes are referred to as nodes that do not split. The term "subtree" can refer to a branch of the decision tree. The selection of attributes in decision trees is based on a number of factors. Entropy, Information Gain, Gini Index, Chi-Square, etc. are a few of these criteria. Information gain evaluates how well an attribute distinguishes training examples, whereas entropy examines the randomness in the information. The cost function used to assess splits in the dataset is referred to as the Gini Index. The link between parent nodes and child nodes is examined using Chi-Square.

6. Random Forest

   The Random Forest [3] technique constructs forests from a collection of decision trees and is used in supervised learning. It is based on bootstrap aggregation to construct forests and is employed for solving regression and classification issues. The method known as "bagging" or "bootstrap aggregation" selects a subset of data at random from the entire dataset. In this case, we use a technique known as row sampling to generate new samples from the original dataset and build our models from those. Bootstrap is shorthand for "sampling with replacement in rows." Each model is then trained individually based on this sampling, yielding results. The ultimate outcome is derived by summing the ratings of all models, which is decided by a simple majority. Aggregation is the process of merging votes into a single outcome that can be used to produce a decision via consensus. The algorithm's final result is determined by averaging the forecasts of several trees, or "decision trees." In order to get a more accurate result, more trees should be used. The Random Forest algorithm deals with the overfitting issue and missing data by enhancing precision.

7. Bagging

   It's a parallel ensemble learning approach that makes use of the autonomy of the individual learners. Typically, the classifiers at the base level are consistent with one another. The training data set is partitioned into many samples by this technique, which is also known as Bootstrap AGGregatING. Then, a decision tree is "fit" to each sample. It is common practise to use many training data sets, each of which is used to create a unique decision tree. Therefore, several decision trees will arrive to dissimilar conclusions. For classification issues, each model's output class is counted as a vote, and the model with the most votes is used to determine the final output class. The training dataset is used to randomly select the samples. During the training phase, the number of decision trees can be set. The final ensemble model significantly reduces the huge variation produced by the individual decision trees. The process of bagging an-ensemble consists of the following steps:

(a) Multiple subsets of the original dataset are created (known as bootstrap samples).

(b) A base learner, typically one Decision Tree per sample, is used to train each subset of the sample.

(c) The weak learners operate concurrently and independently predict the test data.

(d) On the basis of the weak learners' majority vote, the final categorization (malware, benign) is determined.

8. Neural Network

The neural network is a set of algorithms that mimic the functioning of the human brain to identify patterns and relationships in data. This information can then be utilized to make inferences about the data. The term "neural network" refers to both biological and artificial systems consisting of interconnected neurons.

3.4 Un-Supervised Learning Models

Unsupervised machine learning is a type of machine learning where the model is not provided with labeled training examples. Instead, the model is given a dataset and is expected to discover patterns and relationships within the data without any guidance. To validate our data labeling, we used unsupervised models. The workflow of unsupervised machine learning models is represented in figure 3.4. The dataset contains different human vitals like heart rate, body temperature, blood pressure including both systolic and diastolic, blood oxygen saturation level, and respiration rate. After data acquisition, data pre processig is required. For example, the removal of unusual data from raw data. After that Principle Component Analysis (PCA) is used to reduce the dimensions of dataset into 2 values. Unsupervised machine learning models made 2 clusters of dataset. One cluster represent healthy data while other one represented unhealthy data.

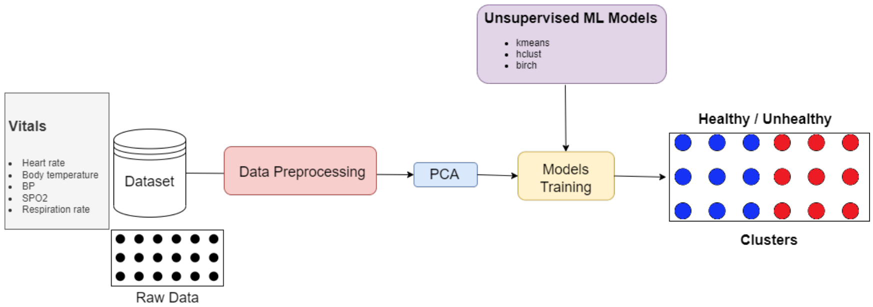


Figure 3.4: Unsupervised Machine Learning Models Workflow

The following is a list of the employed unsupervised learning algorithms:

1. Kmeans

   K-Means clustering is one of the most common clustering methods since it is one of the simplest clustering methods, making it easy to grasp and simple to apply in code. The following mathematical formula (equation: 3.4.1) represents its definition.

$$\underset{C_1,\ldots,C_K}{\text{minimize}}\left\{\sum_{k=1}^{K} W\left(C_k\right)\right\} \tag{3.4.1}$$

   K is the number of all clusters, while C represents each individual cluster. Our goal is to minimize W, which is the measure of within-cluster variation.

$$W\left(C_k\right)=\frac{1}{\left|C_k\right|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}\left(x_{ij}-x_{i'j}\right)^2 \tag{3.4.2}$$

   The most commonly used definition of within-cluster variation is squared Euclidean distance, which is represented by the equation mentioned above. This results in the following form of K-Means clustering, with W being replaced by the euclidean distance formula (equation: 3.4.3).

$$\underset{C_1,\ldots,C_K}{\text{minimize}} \left\{ \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} \left(x_{ij} - x_{i'j}\right)^2 \right\} \tag{3.4.3}$$

2. Hclust

   Hierarchical clustering is an unsupervised learning method used for grouping data points into clusters. The method makes clusters by calculating the variations in the data. Unsupervised learning is a technique that does not require the training of a model or a target variable. Instead, it provides the ability to analyze any data and understand the relationships between data points. The agglomerative clustering method is a bottom-up approach to hierarchical clustering, which starts by treating each data point as a separate cluster. The process continues by merging the closest clusters together until all of the data points are grouped into a single, large cluster.

3. Birch The acronym BIRCH stands for Balanced Iterative Reducing and Clustering Making Use of Hierarchies. This is a clustering method that is able to cluster large datasets by creating a summarized and compact representation of the data while preserving important information. The clustering process is then applied to this reduced summary instead of the entire dataset. By producing a summary of the dataset that may be used by the other clustering method, BIRCH is frequently used to supplement other clustering algorithms. This process takes place in the context of data mining. However, there is a significant limitation with BIRCH, and that is that it can only analyse metric characteristics. Any property the values of which may be represented in Euclidean space is considered to be a metric attribute; in other words, there should not be any categorical attributes present. The first step to using BIRCH is to have a clear understanding of Clustering Feature and Tree Clustering Feature. BIRCH breaks down large datasets into smaller and more dense sections known as Clustering Feature (CF) entries. Each CF entry is defined as an ordered triple, consisting of the number of data points within the cluster (N), the linear sum of these data points (LS), and the squared sum of these data points (SS). It's possible for a CF entry to be composed of parts from several different CF entries. CF Family Tree The actual compact representation, which is what we have been discussing up to this point, is the CF tree. A tree that has sub-clusters included within each of its leaf nodes is known as a CF tree. Every

entry in a CF tree has a pointer that leads to a child node, as well as its own CF entry that is a summation of the CF entries contained in its child nodes. Each leaf node has a maximum capacity for the amount of entries it can contain. The term "threshold" refers to this highest possible number. We shall acquire further knowledge regarding the nature of this threshold value. The BIRCH algorithm's parameters are as follows:

- threshold : A sub-cluster in the leaf node of the CF tree can retain a maximum number of data points known as threshold.
- branching_factor : This parameter specifies the maximum amount of CF sub-clusters per node.
- Then_clusters : The amount of clusters that will be returned once the BIRCH algorithm has finished, or how many clusters there are after the very last clustering step. The final clustering step is skipped and intermediate clusters are returned if set to None.

Chapter 4

Results

In this chapter, the results of the research regarding the machine learning module are discussed in detail. The various algorithms applied are discussed and a comparative analysis is presented based on the derived results. The debate is backed up by conventional performance measurements as well as the consequent accuracy.

4.1 Evaluation Metrics

The evaluation metrics are used to determine the model's effectiveness or quality. Accuracy, precision, recall, Fl-score, and confusion metrics are the evaluation metrics utilised in this study work to evaluate the machine learning learning models.

- Accuracy:

  It is the proportion of samples that were correctly categorised compared to the proportion that were incorrectly classified. Equation 4.1.1 is used to determine the value.

  $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1.1}$$

  Where,

  The term "true positive" refers to the number of tests that accurately identify positive samples.

  True Negative is the percentage of samples that were accurately identified as negative.

The term "false positive" describes the occurrence of an incorrectly positive result after a series of "negative" samples has been processed.

The term "false negative" describes the occurrence in which a positive sample is incorrectly labelled as having a negative result.

- Precision:

  It is the proportion of samples labelled as positive overall to samples that are actually positive. Equation 4.1.2 is employed to calculate it.

$$Precision = \frac{TP}{TP + FP} \tag{4.1.2}$$

- Recall:

  It shows the proportion of TP samples among all positive samples, as determined by equation 4.1.3

$$Recall = \frac{TP}{TP + FN} \tag{4.1.3}$$

- Fl-score:

  It is the balance between the precision and recall which is computed by using equation 4.1.4

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.1.4}$$

- Confusion Matrix:

  It's a table that gives details of classifier performance as shown in figure 4.1. By visualising the confusion matrix, the other performance metrics, including precision, recall, and accuracy, may be determined.

4.2 Supervised Machine Learning models

A brief detail of the tested models and their results are described below:

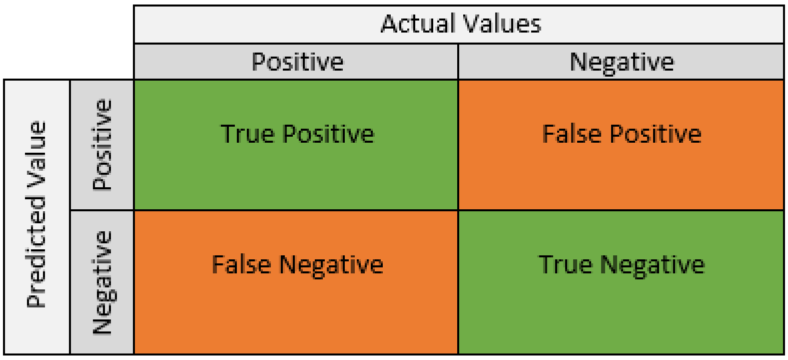


Figure 4.1: Confusion Matrix

4.2.1 Logistical Regression

Logistic regression is another type of statistical tool that is used in machine learning. It is the preferred strategy for problems involving binary categorization. We trained a model using logistic regression on our dataset which achieved an accuracy of 90 % and an MSE of 0.1014 on the test data. The confusion matrix of the Logistical Regression model is shown below in figure 4.2.

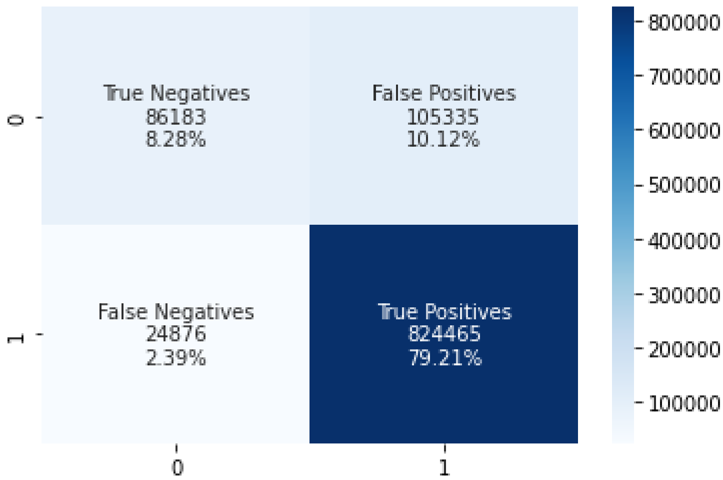


Figure 4.2: Confusion Matrix of Logistical Regression

4.2.2 Naïve Bayes

The naive Bayes classifiers are a type of straightforward probabilistic classifiers used in data science. We also trained a model with this algorithm. The model achieved an accuracy of 78% and an MSE of 0.22 . The confusion matrix of the naive bayes is shown in figure 4.3.

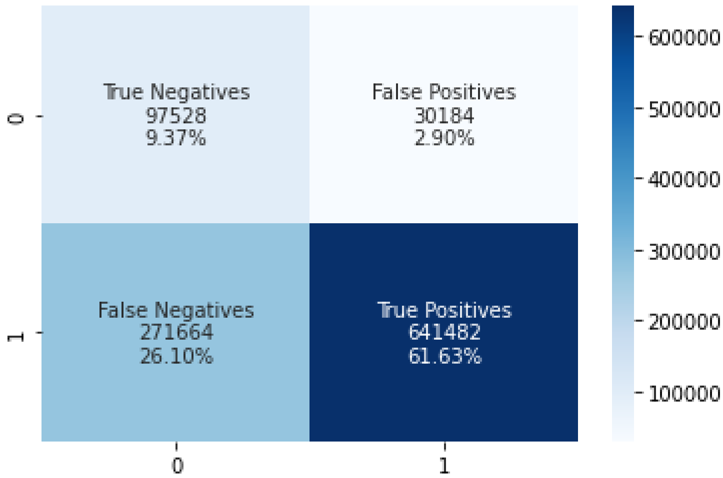


Figure 4.3: Confusion Matrix of Naïve Bayes

4.2.3 K-Nearest Neighbors (KNN)

A KNN approach is really straightforward, although it can be utilised for complex applications and obscure dataset distributions. The KNNs algorithm achieved the best performance in our experiments. We trained a model with K=17 and achieved an accuracy of 93% and an MSE of 0.062 . The confusion matrix of the KNN is shown in figure 4.4.

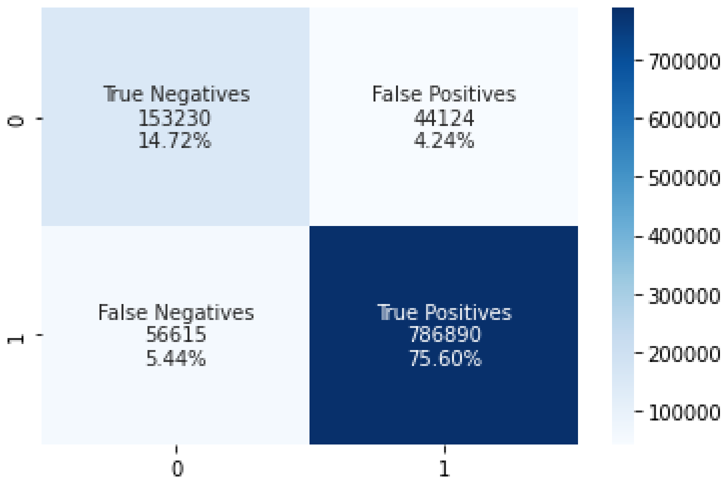


Figure 4.4: Confusion Matrix of KNN

4.2.4 Support Vector Machines (SVM)

Support Vector Machines (SVM) are a type of supervised machine learning algorithm that perform both classification and regression tasks. The learning algorithms associated with SVM analyze the data and classify or predict the outcome based on the labeled training data. We trained a SVM model on our training data achieving an accuracy of 91% and an MSE of 0.041. SVMs are very strong binary classifiers, SVM's confusion matrix is shown in figure 4.5.

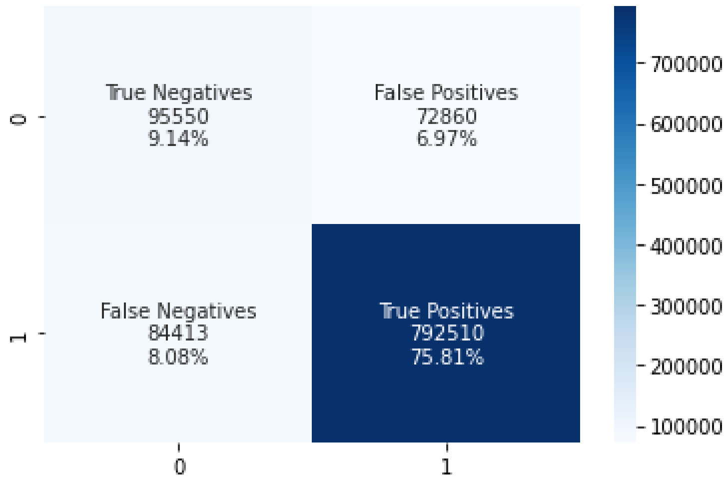


Figure 4.5: Confusion Matrix of SVM

4.2.5 Decisions Trees

A decision tree are used in machine learning to formally and visually describe decisions and decision-making. As implied by the name, it employs a decision-tree model. We trained a decision trees model on our training data achieving an accuracy of 92% and an MSE of 0.089 . Decision Tree's confusion matrix is shown in figure 4.6.

4.2.6 Random Forest

An ensemble learning classification method is random forests. During training, a large number of decision trees are created using this technique, and the class that is produced is the median of the classes or the mean/average forecast of all the trees. When we trained a random forest classifier with 130 estimators, we achieved an accuracy of 95% and an MSE of 0.0309 . The confusion matrix of the applied model is shown below in

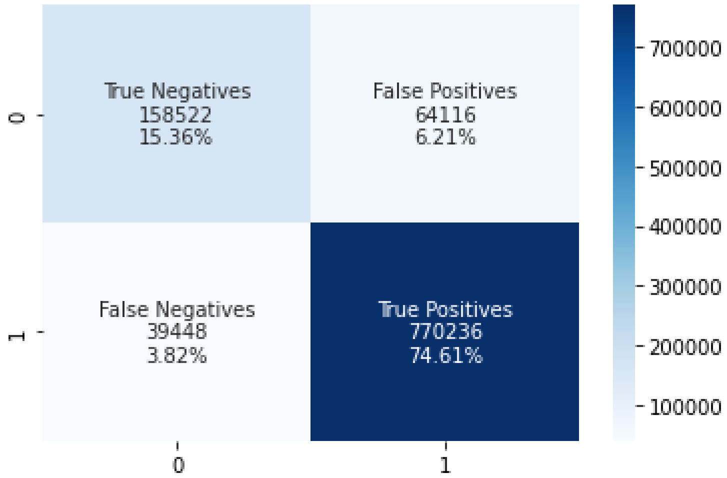


Figure 4.6: Confusion Matrix of Decisions Trees

figure 4.7.

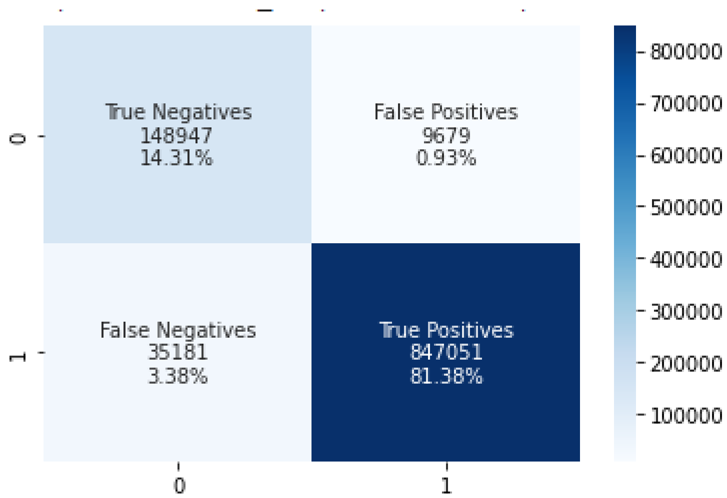


Figure 4.7: Confusion Matrix of Random Forest

4.2.7 Bagging

Bagging, also known as bootstrap aggregating, is a machine learning ensemble meta-algorithm that aims to improve the precision and reliability of machine learning algorithms used in statistical classification and regression. It creates multiple models and combines their results to achieve a more accurate and stable prediction. As a result, overfitting is prevented and variation is reduced. We trained a bagging model which achieved an accuracy of 94 % and an MSE of 0.9 .

4.2.8 Neural Network

Neural networks are systems made up of nodes that are linked together and work like neurons in the brain. Using algorithms, they may be able to find hidden patterns and relationships in raw data, group and classify it, and get better at it over time. We were able to train a neural network with three hidden layers and attain an accuracy of 89.99 percent.

4.3 Comparative Analysis of Machine Learning Models

These supervised machine learning algorithms have been evaluated using various performance metrics, such as accuracy, precision, recall, f1 score, and error rate. The results of these evaluations provide insights into the effectiveness and reliability of the algorithms. Details are mentioned below:

4.3.1 Model Accuracy

The model's corrected percentage of predictions is determined by calculating the accuracy of the machine learning algorithm. Figure 4.8 displays the accuracy of various machine learning models.

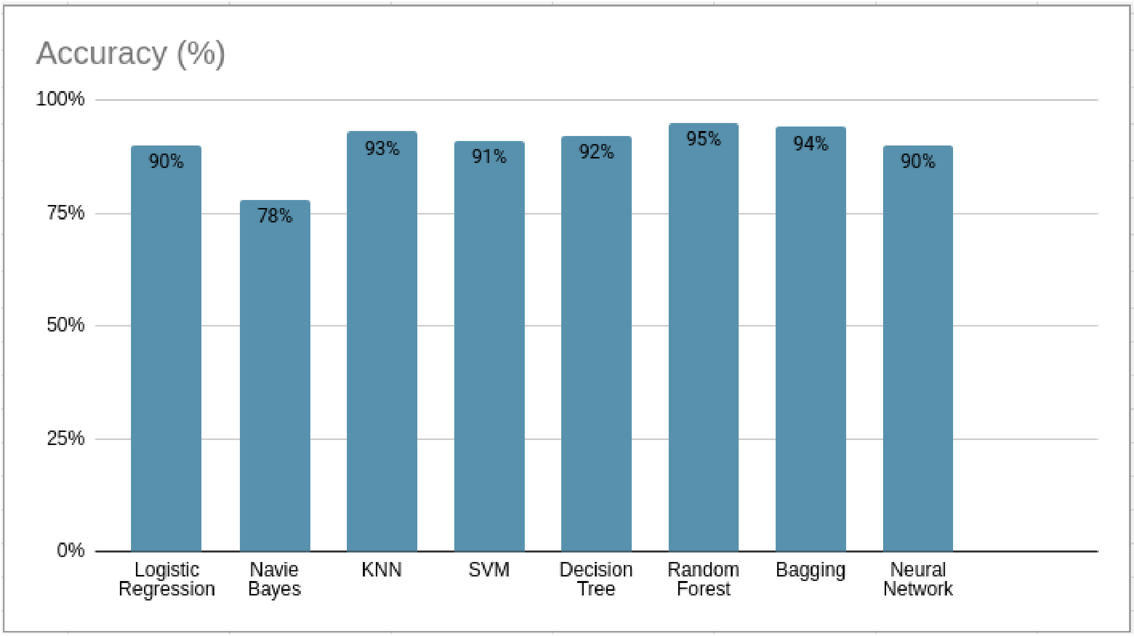


Figure 4.8: ML Model Accuracy

By comparing the results of several cross-validations, we find that random forest achieves the maximum accuracy of 95%. Cross-validation is an iterative process in which the data is repeatedly divided into two parts: a training set and a test set. The machine learning model is trained on the training set and evaluated on the test set. This process is repeated multiple times with different divisions of the data, and the results are averaged to obtain the final validation measures. Conversely, Bayes' algorithm is only 78% accurate. It is because of the relative nature of the mimic-III dataset that this classifier's performance suffers; it treats the input features as if they were completely unrelated.

4.3.2 Model Errors

Mean Square Error (MSE) and R-Squared Error are two important metrics used in the evaluation of machine learning models. MSE measures the average of the squared differences between the predicted and actual values, while R-Squared Error provides a measure of the amount of variability in the target variable that is explained by the model. These metrics give a clear summary of the performance of a machine learning model, as depicted in figure 4.9 Since the Mimic III dataset utilised in this study has imbalanced classes, we may conclude from figure 4.9 that the Naive Bayes method has the most errors because its functionality did not match the data set, whereas Random Forest has the fewest errors since it is sensitive to class imbalance.

4.3.3 Models Metrics

Machine learning algorithms are assessed using a variety of execution metrics.In order to assess ML models, it is crucial to consider their Precision, Recall, and F1 scores.

The figure 4.10 depicts the performance metrics of ML models.This illustrates that Random Forest has superior performance compared to Naive Bayes. Nevertheless, KNN and Decision Tree perform similarly to Random Forest.

4.3.4 Confusion Matrix

The predictions of the ML model are evaluated using a confusion matrix. The Confusion matrix of applied machine learning models is displayed in figure 4.11. It is clear that Naive Bayes makes more inaccurate predictions than other models, which reduces its

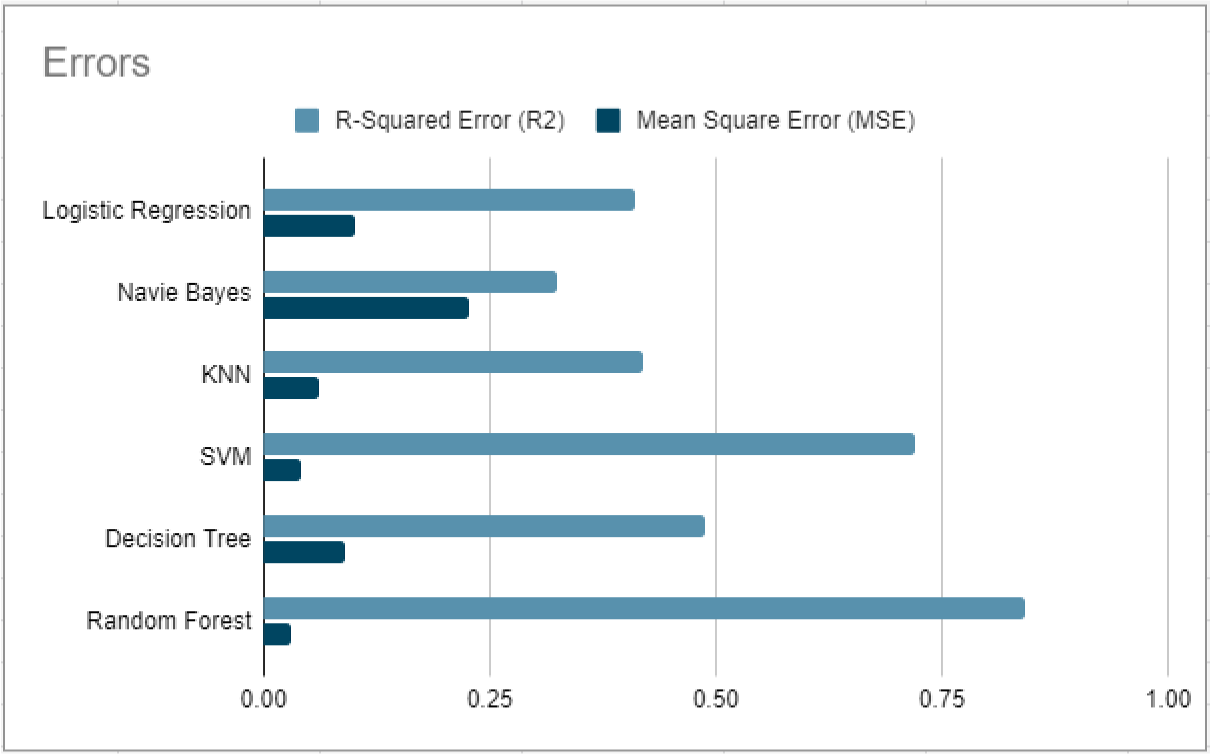


Figure 4.9: ML Model Errors

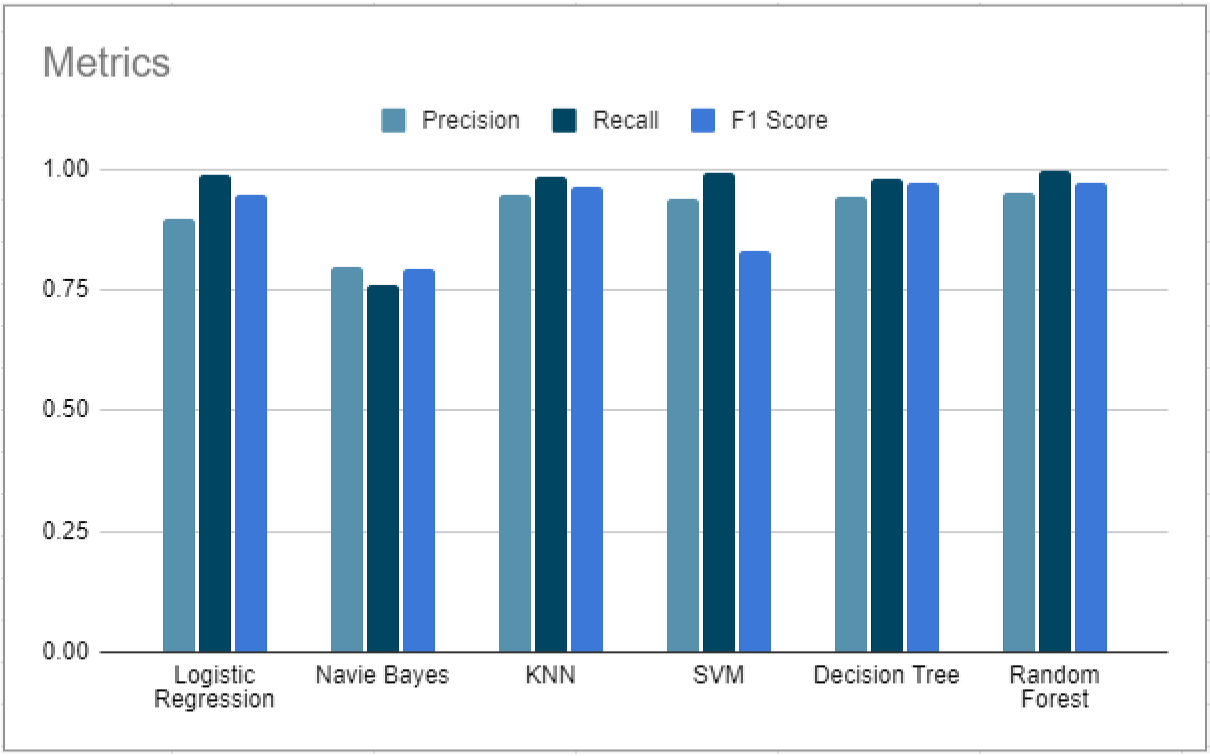


Figure 4.10: ML Model metrics

accuracy to 78 percent. With the exception of Naive Bayes, the other algorithms' True Positive results range from 70 to 81 percent, which is regarded as good performance.

The True negative values are another standout feature of the confusion matrices. The unbalanced nature of the dataset may be to blame for the low numbers. Given that the Random Forest algorithm performs very well on unbalanced classification issues, it is clear from the findings that the model's error rates for False Positive and False Negative predictions, respectively, are quite low at 0.93 and 3.38 percent, respectively. This explains why Random Forest performed generally better on the Mimic-III dataset than other ML models.

Figure 4.11: ML Model Confusion Matrix

According to the above information, Random Forest is the most effective algorithm for detecting cardiac health anomalies.

4.4 Un Supervised Machine Learning models

Clustering is a key task in unsupervised machine learning. This task's objective is to group instances of a given dataset into various clusters according to their shared traits.

For the K-Means, hclust, and birch clustering models, we produced a 2D Principal Component Analysis (PCA) plot in this scenario. Dimensionality reduction, also known as data projection to a lower-dimensional space, is a method that uses PCA to keep the majority of the variation. The dataset's original 6 characteristics have been condensed to 2 principle components after applying PCA to it.

These unsupervised ML methods are used to create clusters, which are then discussed following:

4.4.1 Kmeans

K-Means clustering is a popular and simple clustering method that is simple to understand and implement in code. The cluster of the applied model is depicted in figure 4.12.

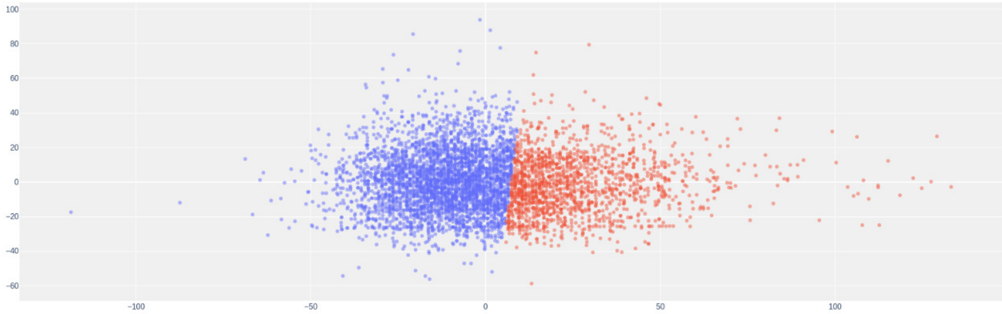


Figure 4.12: K Means Cluster

4.4.2 hclust

Hclust is also known as Hierarchical Clustering. It made clusters by measuring the dissimilarities between the data. The cluster of the applied model is shown below in figure 4.13.

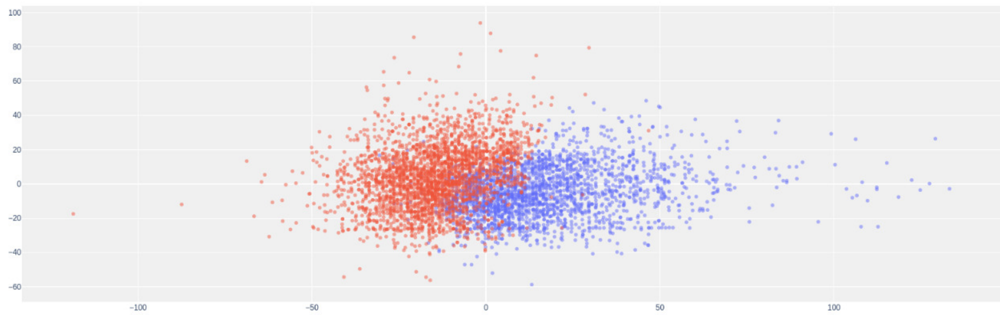


Figure 4.13: Hclust Cluster

4.4.3 BIRCH cluster

BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies) is a clustering algorithm designed for handling large datasets. It groups similar items together by creating a summarized representation of the dataset that retains as much information as possible, making it more efficient and scalable for processing extremely large data sets. The cluster of the applied model is depicted in figure 4.14.

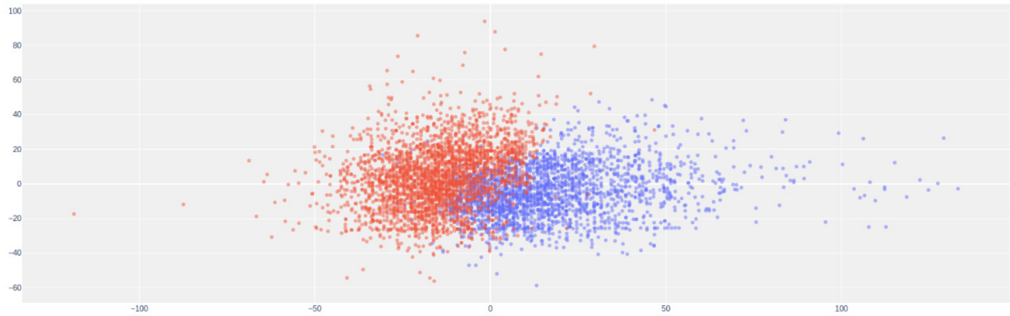


Figure 4.14: Birch cluster

According to the above analysis, it is clearly visible that clusters of K means is more clear than hclust and BIRCH. As there is no overlapping in between clusters.

Chapter 5

Data Visualization

The web application was created to visualize the information for both patients and physicians and to facilitate communication between them. It provides a platform for effectively presenting the information to both parties and enabling them to communicate and exchange information. The purpose of the application was to give the doctor remote access to patient data that would not have otherwise been possible. The website's goal is to make it simple and convenient for patients and doctors to exchange information, as well as to offer any additional information that may be required in addition to the real-time display of physical and physiological parameters and predictions from the machine learning model for more accurate and timely diagnosis.

The web application offers several key features, including data security, alert notifications, user-friendly interfaces, and real-time data display. These features provide a secure and efficient platform for visualizing and exchanging information between patients and physicians. Admin added patients and doctors in the system so that they could access the website. The website shows the user's personal information, diagnosis comments, and vitals data.

All of the page designs were inspired by already existing applications with a similar objective. The website maintains unity by utilizing constant designs and color schemes. Each active and recognisable link on the website is responsive.

The architecture diagram of a web portal is shown in figure 5.1. React JS is used for the development of graphical user interfaces. React app call the Hyper Text Transfer Protocol (HTTP) Rest API. Node JS is used to make an HTTP Rest API and write the business logic of the system. For data storage, a PostgreSQL database is used.

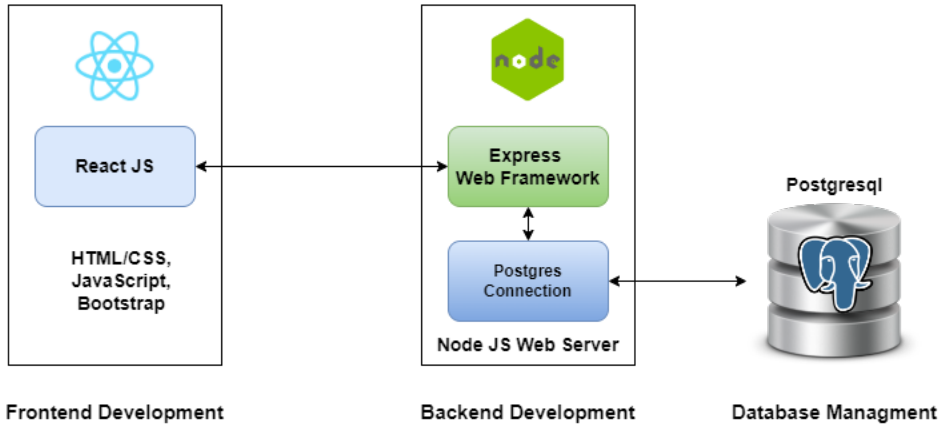


Figure 5.1: Web portal architecture diagram

5.1 Technologies

5.1.1 Client and Server Architecture:

The client-server architecture is a distributed architecture used in web portal development. In this architecture, multiple clients send requests for services to a centralized server, which provides the requested services to the clients. This architecture enables efficient and scalable management of resources and data. The main processing and business logic are handled on the server side. The client sends a request for any service and in response, the server sends the result back to the client. The purpose of using this architecture is to deploy the ML model on the server-side so that clients send data to the server and data processing is easily handled on the server-side.

5.1.2 Server-side:

For server development, Node JS is mainly used. Javascript is mostly used in Node JS. For machine learning model deployment, python is used. We have to install pandas, NumPy, scikit-image, TensorFlow, PyTorch, and pillow libraries on the server-side for detection of anomaly in human vitals.

5.1.3 Client-side:

React JS is used for the development of the client-side application. It is an open-source JavaScript library for constructing UI components-based front-end user interfaces. Graphical User Interface is made for clients to detect anomaly in human vitals on the website, and the vitals data is sent to the server for further processing and detection of anomalies in human vitals.

5.1.4 Database:

PostgreSQL is an relational database. That is also known as postgres. It support both SQL and JSON queries. Postgres support mostly popular languages like Python, Java, C#, C++, C, JavaSCript, Go, Ruby etc.

PgAdmin 4 is an application to support PostgreSQL database. It provide Graphical User Interface (GUI) to end user. User can easily connect postgres databse with pgadmin to view data as well as to run any query like create, read, update, and delete.

5.1.5 Node JS:

Node.js is an open-source, server-side platform built on Google Chrome's JavaScript V8 Engine. It is commonly used for building back-end APIs and is known for its event-driven, non-blocking I/O model, making it efficient for real-time applications. It handle API requests. Main business logic is written in node js.

5.1.6 React JS:

React JS is frontend open source javascript library for developing user interfaces. It is used to develop signle page application. Babel is an java compiler that is used to translate programming language into JavaScript. In react it converts JavaScript XML (JSX) into Javascript. JSX is an HTML or XML like extensions.

For frontend development, React Redux is used. Redux is an open-source JS framework used in react js for state management. It enables React components to read the data from a Redux store and send Action to store updated data.

5.1.7 AWS EC2:

Amazon Elastic Computing Cloud provides a platform for cloud computing that rents out virtual computers to users to run computer applications. AWS provides scalable deployment of applications. Furthermore, it also provides elasticity. Users have the option to create, launch and terminate AWS instances when needed and pay only by use. EC2 is used to deploy web applications on AWS.

The web portal offers end users a wide range of capabilities, such as data privacy, alerts for emergency warnings, and real-time data display in tabular and graphical style, as illustrated in figure 5.3 and figure 5.2 respectively. This website provides remote patient monitoring and management services. Depending on their security permissions, each user can access just their relevant data.

Figure 5.2: Graphical Data Visualization

There are four primary user roles, which are as follows:

- Admin

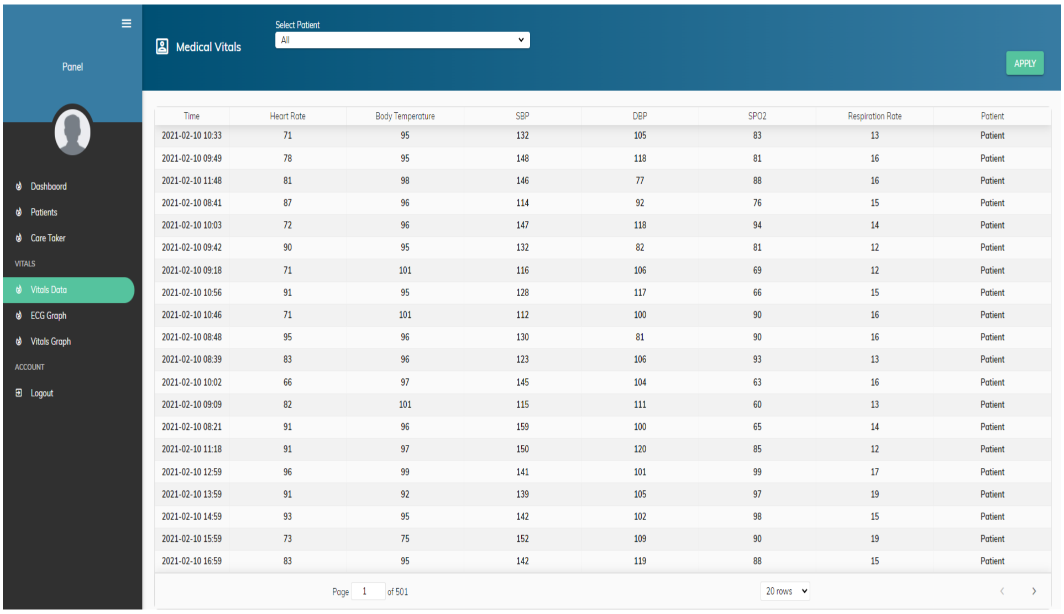
  Admin can manage all other users of system doctors, patients and caretakers. So that other users can access the system according to their role.

- Doctor

  Doctor is able to only view his patients information, his/her medical vitals. Historical data of patient's vitals can also be viewed in tabular format. Daily and weekly graphs of vitals can also be viewed for analysis.

- Patient

  Patient can view his latest human vitals. He/She can view his/her own historical data and in graphical format also.



Time	Heart Rate	Body Temperature	SBP	DBP	SPO2	Respiration Rate	Patient
2021-02-10 10:33	71	95	132	105	83	13	Patient
2021-02-10 09:49	78	95	148	118	81	16	Patient
2021-02-10 11:48	81	98	146	77	88	16	Patient
2021-02-10 08:41	87	96	114	92	76	15	Patient
2021-02-10 10:03	72	96	147	118	94	14	Patient
2021-02-10 09:42	90	95	132	82	81	12	Patient
2021-02-10 09:18	71	101	116	106	69	12	Patient
2021-02-10 10:56	91	95	128	117	66	15	Patient
2021-02-10 10:46	71	101	112	100	90	16	Patient
2021-02-10 08:48	95	96	130	81	90	16	Patient
2021-02-10 08:39	83	96	123	106	93	13	Patient
2021-02-10 10:02	66	97	145	104	63	16	Patient
2021-02-10 09:09	82	101	115	111	60	13	Patient
2021-02-10 08:21	91	96	159	100	65	14	Patient
2021-02-10 11:18	91	97	150	120	85	12	Patient
2021-02-10 12:59	96	99	141	101	99	17	Patient
2021-02-10 13:59	91	92	139	105	97	19	Patient
2021-02-10 14:59	93	95	142	102	98	15	Patient
2021-02-10 15:59	73	75	152	109	90	19	Patient
2021-02-10 16:59	83	95	142	119	88	15	Patient

Figure 5.3: Recorded Human Vitals

- Caretaker

  Caretaker can view their patient's latest bio vitals. They can view their patient historical data and in graphical format also.

The administrator holds the power to manage user access to the web portal by adding or removing users. They are responsible for the operation and maintenance of the online platform. Due to the confidential nature of medical information, data security and privacy are a high priority. To ensure security, the portal includes two-factor authentication in its design. Only authorised personnel are allowed access to the recorded data.

Chapter 6

Conclusion

In this study, a cardiac health monitoring system based on the Internet of Medical Things (IoMT) has been introduced. The system uses machine learning models to identify unusual patterns in medical vitals related to heart health. We have analyzed eight different supervised classification machine learning techniques for anomaly detection. In these trained models, the accuracy of Random Forest is 95%. To cross validation of data labeling, we used three unsupervised learning models. In which kmean clusters performed very well. Finally for remote health monitoring, Web portal is developed in latest technologies so that doctor can examine patient health remotely.

6.1 Future Work

We want to build a wearable sensor vest that focuses on monitoring cardiovascular health and has an interface to an Edge device as the next step in our research. This study incorporates a cardiac health monitoring system that leverages the Internet of Medical Things (IoMT) and utilizes machine learning models for anomaly detection. After evaluating and comparing eight different supervised machine learning techniques, the most effective model will be implemented on an edge device for performance evaluation. The edge device will be linked to the AWS cloud for control and cloud computing through AWS IoT-Greengrass, and patient medical data will be stored in the AWS Relational Database Service (RDS) for secure access by both patients and physicians through a web-based portal. We will strengthen this research by employing our proposed health monitoring device to gather a local data set with the aid of partner health institutions.

This data will be subject to all requirements for medical records. This includes information on data protection and confidentiality, patient consent forms, etc. Finally, it will be given a designation by medical experts. The system's usability and functionality will both improve as a result of these suggested improvements. Additionally, it will provide a precise validation of our suggested system. We are certain that this technology would significantly improve Pakistan's healthcare system. This facility will provide doctors and patients with out-of-hospital surveillance, which is critical in the present COVID-19 pandemic age. In addition, this framework will provide distant patients with home-based access to reliable, high-quality health monitoring system. Cardiologists will benefit from the machine learning classifiers' decision-support tools, which will also hasten the creation of patient treatment plans. The suggested system would be developed in accordance with industry standards and will have a high degree of communication efficiency, enabling the system to react rapidly in the event that medical professionals and caretakers receive emergency alerts. For patients who are seriously unwell, these cautions may be helpful.

Bibliography

[1] J. Ross Quinlan. “Induction of decision trees”. In: *Machine learning* 1.1 (1986), pp. 81–106.

[2] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine learning* 20.3 (1995), pp. 273–297.

[3] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[4] Joseph J Oresko et al. “A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing”. In: *IEEE Transactions on Information Technology in Biomedicine* 14.3 (2010), pp. 734–740.

[5] KR Darshan and KR Anandakumar. “A comprehensive review on usage of Internet of Things (IoT) in healthcare system”. In: *2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT).* IEEE. 2015, pp. 132–136.

[6] Desy Dwi Purnomo et al. “Design of pulse rate and body temperature monitoring system with Arduino via wifi and android-based gadget”. In: *International Journal of Technology and Engineering Studies* 2.5 (2016), pp. 140–148.

[7] Arijit Ukil et al. “IoT healthcare analytics: The importance of anomaly detection”. In: *2016 IEEE 30th international conference on advanced information networking and applications (AINA).* IEEE. 2016, pp. 994–997.

[8] Sruthy Orozhiyathumana Agnisarman et al. “Lessons learned from the usability assessment of home-based telemedicine systems”. In: *Applied ergonomics* 58 (2017), pp. 424–434.

[9] R Ani et al. “Iot based patient monitoring and diagnostic prediction tool using ensemble classifier”. In: *2017 International conference on advances in computing, communications and informatics (ICACCI).* IEEE. 2017, pp. 1588–1593.

[10] MR Desai and Sushma Toravi. "A smart sensor interface for smart homes and heart beat monitoring using WSN in IoT environment". In: *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. IEEE. 2017, pp. 74–77.

[11] S Pradeep Kumar et al. "Smart health monitoring system of patient through IoT". In: *2017 international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC)*. IEEE. 2017, pp. 551–556.

[12] Francisco Mugica et al. "A model for continuous monitoring of patients with major depression in short and long term periods". In: *Technology and Health Care* 25.3 (2017), pp. 487–511.

[13] Melisa Pereira and Kamath K Nagapriya. "A novel IoT based health monitoring system using LPC2129". In: *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE. 2017, pp. 564–568.

[14] Uttara Gogate and Jagdish Bakal. "Healthcare monitoring system based on wireless sensor network for cardiac patients". In: *Biomedical & Pharmacology Journal* 11.3 (2018), p. 1681.

[15] Sandeep Pirbhulal et al. "A medical-IoT based framework for eHealth care". In: *2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI)*. IEEE. 2018, pp. 1–4.

[16] Fatemeh Rezaeibagha and Yi Mu. "Practical and secure telemedicine systems for user mobility". In: *Journal of biomedical informatics* 78 (2018), pp. 24–32.

[17] Mehmet Taştan. "IoT based wearable smart health monitoring system". In: *Celal Bayar University Journal of Science* 14.3 (2018), pp. 343–350.

[18] Jie Wan et al. "Wearable IoT enabled real-time health monitoring system". In: *EURASIP Journal on Wireless Communications and Networking* 2018.1 (2018), pp. 1–10.

[19] Sidong Wei, Xuejiao Zhao, and Chunyan Miao. "A comprehensive exploration to the machine learning techniques for diabetes identification". In: *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE. 2018, pp. 291–295.

[20] DM Jeya Priyadharsan et al. "Patient health monitoring using IoT with machine learning". In: *International Research Journal of Engineering and Technology (IRJET)* 6.03 (2019).

[21] Protap Kumar Saha, Nazmus Sakib Patwary, and Ifthakhar Ahmed. "A widespread study of diabetes prediction using several machine learning techniques". In: *2019 22nd International Conference on Computer and Information Technology (ICCIT)*. IEEE. 2019, pp. 1–5.

[22] K Durga Saranya et al. "IoT-based health monitoring system using beaglebone black with optical sensor". In: *Journal of Optical Communications* (2019).

[23] Christoph Stach and Bernhard Mitschang. "ECHOES: A Fail-Safe, Conflict Handling, and Scalable Data Management Mechanism for the Internet of Things". In: *European Conference on Advances in Databases and Information Systems*. Springer. 2019, pp. 373–389.

[24] Heena Varshney, Ali S Allahloh, and Mohammad Sarfraz. "IoT Based eHealth Management System Using Arduino and Google Cloud Firestore". In: *2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE. 2019, pp. 1–6.

[25] Khin Thet Wai, Nyan Phyo Aung, and Lwin Lwin Htay. "Internet of things (IoT) based healthcare monitoring system using NodeMCU and Arduino UNO". In: *Published in International Journal of Trend in Scientific Research and Development (ijtsrd)* 3.5 (2019), pp. 755–759.

[26] A Athira, TD Devika, KR Varsha, et al. "Design and development of IOT based multi-parameter patient monitoring system". In: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE. 2020, pp. 862–866.

[27] Saurabh Bhattacharya and Manju Pandey. "Significance of IoT in India's E-medical framework: A study". In: *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*. IEEE. 2020, pp. 321–324.

[28] Saiteja Prasad Chatrati et al. "Smart home health monitoring system for predicting type 2 diabetes and hypertension". In: *Journal of King Saud University-Computer and Information Sciences* (2020).

[29] Stephen Dewanto, Michelle Alexandra, and Nico Surantha. "Heart Rate Monitoring with Smart Wearables using Edge Computing". In: *International Journal of Advanced Computer Science and Applications* 11.3 (2020).

[30] Chunyan Guo et al. "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform". In: *IEEE Access* 8 (2020), pp. 59247–59256. DOI: 10.1109/ACCESS.2020.2981159.

[31] Chunyan Guo et al. "Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform". In: *IEEE Access* 8 (2020), pp. 59247–59256.

[32] Al-Zadid Sultan Bin Habib and Tanpia Tasnim. "An ensemble hard voting model for cardiovascular disease prediction". In: *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE. 2020, pp. 1–6.

[33] Saqib Hakak et al. "A framework for edge-assisted healthcare data analytics using federated learning". In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. 2020, pp. 3423–3427.

[34] Md Islam, Ashikur Rahaman, et al. "Development of smart healthcare monitoring system in IoT environment". In: *SN computer science* 1.3 (2020), pp. 1–11.

[35] Yuanyuan Pan et al. "Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform". In: *IEEE Access* 8 (2020), pp. 189503–189512.

[36] Honey Pandey and S Prabha. "Smart health monitoring system using IOT and machine learning techniques". In: *2020 sixth international conference on bio signals, images, and instrumentation (ICBSII)*. IEEE. 2020, pp. 1–4.

[37] Md Raseduzzaman Ruman et al. "IoT based emergency health monitoring system". In: *2020 International Conference on Industry 4.0 Technology (I4Tech)*. IEEE. 2020, pp. 159–162.

[38] Shreshth Tuli et al. "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments". In: *Future Generation Computer Systems* 104 (2020), pp. 187–200.

[39] T Vairam et al. "Real-Time Cardiovascular Health Monitoring System Using IoT and Machine Learning Algorithms: Survey". In: *Securing IoT and Big Data*. CRC Press, 2020, pp. 145–162.

[40] Prajoona Valsalan, Tariq Ahmed Barham Baomar, and Ali Hussain Omar Baabood. "IoT based health monitoring system". In: *Journal of critical reviews* 7.4 (2020), pp. 739–743.

[41] Ionel Zagan et al. "Design, fabrication, and testing of an IoT healthcare cardiac monitoring device". In: *Computers* 9.1 (2020), p. 15.

[42] Tommaso Zoppi, Andrea Ceccarelli, and Andrea Bondavalli. "Into the unknown: Unsupervised machine learning algorithms for anomaly-based intrusion detection". In: *2020 50th Annual IEEE-IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S)*. IEEE. 2020, pp. 81–81.

[43] G Bhanuteja et al. "CAMISA: An AI Solution for COVID-19". In: *2021 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*. IEEE. 2021, pp. 216–222.

[44] Essam H Houssein et al. "Integration of internet of things and cloud computing for cardiac health recognition". In: *Metaheuristics in Machine Learning: Theory and Applications*. Springer, 2021, pp. 645–661.

[45] Md Abdur Rahman and M Shamim Hossain. "An internet-of-medical-things-enabled edge computing framework for tackling COVID-19". In: *IEEE Internet of Things Journal* 8.21 (2021), pp. 15847–15854.

[46] Umara Umar et al. "IoT-based Cardiac Healthcare System for Ubiquitous Healthcare Service". In: July 2021, pp. 1–6. DOI: 10.1109/ICOTEN52080.2021.9493478.

[47] Zarlish Ashfaq et al. "Embedded AI-Based Digi-Healthcare". In: *Applied Sciences* 12.1 (2022), p. 519.

[48] R.M Johnson A.; Tom Pollard. *MIMIC-III Clinical Database (Version 1.4). PhysioNet 2016, 10, C2XW26*.

[49] MedlinePlus. *Vital Signs. Medical Encyclopedia. Available online:* https://medlineplus.gov/ency/article/002341.htm.

[50] International The News. *An estimated 30 to 40,hour due to heart attack*. https://www.thenews.com.pk/print/153465-Coronary-artery-diseases-leading-cause-of-death-in-Pakistan.