

Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System

Shadman Nashif¹, Md. Rakib Raihan², Md. Rasedul Islam², Mohammad Hasan Imam²

¹Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology (RUET), Rajshahi, Bangladesh

²Department of Electrical & Electronic Engineering, American International University-Bangladesh (AIUB), Dhaka, Bangladesh
Email: engineereju.cseruet@gmail.com

How to cite this paper: Nashif, S., Raihan, Md.R., Islam, Md.R. and Imam, M.H. (2018) Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System. *World Journal of Engineering and Technology*, 6, 854-873.
<https://doi.org/10.4236/wjet.2018.64057>

Received: September 23, 2018

Accepted: November 19, 2018

Published: November 22, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Cardiovascular diseases are the most common cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate. However, accurate detection of heart diseases in all cases and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. In this study, a tentative design of a cloud-based heart disease prediction system had been proposed to detect impending heart disease using Machine learning techniques. For the accurate detection of the heart disease, an efficient machine learning technique should be used which had been derived from a distinctive analysis among several machine learning algorithms in a Java Based Open Access Data Mining Platform, WEKA. The proposed algorithm was validated using two widely used open-access database, where 10-fold cross-validation is applied in order to analyze the performance of heart disease detection. An accuracy level of 97.53% accuracy was found from the SVM algorithm along with sensitivity and specificity of 97.50% and 94.94% respectively. Moreover, to monitor the heart disease patient round-the-clock by his/her caretaker/doctor, a real-time patient monitoring system was developed and presented using Arduino, capable of sensing some real-time parameters such as body temperature, blood pressure, humidity, heartbeat. The developed system can transmit the recorded data to a central server which are updated every 10 seconds. As a result, the doctors can visualize the patient's real-time sensor data by using the application and start live video streaming if instant medication is required. Another important feature of the proposed system was that as soon as any real-time parameter of the patient exceeds the

threshold, the prescribed doctor is notified at once through GSM technology.

Keywords

Data Mining, Machine Learning, IoT (Internet of Things), Patient Monitoring System, Heart Disease Detection and Prediction

1. Introduction

The heart is a kind of muscular organ which pumps blood into the body and is the central part of the body's cardiovascular system which also contains lungs. Cardiovascular system also comprises a network of blood vessels, for example, veins, arteries, and capillaries. These blood vessels deliver blood all over the body. Abnormalities in normal blood flow from the heart cause several types of heart diseases which are commonly known as cardiovascular diseases (CVD). Heart diseases are the main reasons for death worldwide. According to the survey of the World Health Organization (WHO), 17.5 million total global deaths occur because of heart attacks and strokes. More than 75% of deaths from cardiovascular diseases occur mostly in middle-income and low-income countries. Also, 80% of the deaths that occur due to CVDs are because of stroke and heart attack [1]. Therefore, detection of cardiac abnormalities at the early stage and tools for the prediction of heart diseases can save a lot of life and help doctors to design an effective treatment plan which ultimately reduces the mortality rate due to cardiovascular diseases. Due to the development of advance healthcare systems, lots of patient data are nowadays available (*i.e.* Big Data in Electronic Health Record System) which can be used for designing predictive models for Cardiovascular diseases. Data mining or machine learning is a discovery method for analyzing big data from an assorted perspective and encapsulating it into useful information. "Data Mining is a non-trivial extraction of implicit, previously unknown and potentially useful information about data" [2]. Nowadays, a huge amount of data pertaining to disease diagnosis, patients etc. are generated by healthcare industries. Data mining provides a number of techniques which discover hidden patterns or similarities from data. Therefore, in this paper, a machine learning algorithm is proposed for the implementation of a heart disease prediction system which was validated on two open access heart disease prediction datasets.

Another contribution of this paper is the presentation of a cardiac patient monitoring system using the concept of Internet of Things (IoT) with different physiological signal sensors and Arduino microcontroller. Sensor networks are currently using the Internet of Things (IoT) technology to collect, analyze and passing of information from one node to another. IoT is a recently used rapidly expanding technology, where multiple sensors/data collectors can sense, share information and communicate over a private network, Internet Protocol (IP) or

public networks. The sensors collect the data after a specific time, analyze it and use it to initiate the required action, and provide an intelligent cloud-based network for analysis, planning and decision making. The products that are developed with IoT such as embedded technology, allow to exchange information among each other nodes or the Internet and it was assessed that about 8 to 50 billion devices will be connected by 2020 [3].

2. Background and Objectives

The importance and advantages of the application of Machine learning based heart disease detection and prediction system were discussed in several research findings. The application of artificial intelligence in disease detection system especially the cardiac disease system detection improves the performance of other existing widely used models like models provided by American College of Cardiology/American Heart Association (ACC/AHA) models in CVD detection and prediction [4]. The possibility and related matters of providing advanced services of a human health management system were analyzed by Zhao, Wang, and Nakahira, in 2011 and they had also given a research direction of medical technology on IoT [5]. Many types of health-related sensors and technologies were analyzed by them. They identified some issues which need to be solved. The home monitoring system and decision support system was schemed by Chiu-chisan and Geman in 2014 [6]. This system contributed to home monitoring, diagnosis, medical prescriptions, medical treatment, rehabilitation and development of his patients with Parkinson's disease. Wireless Health Monitoring System (WHMS) has attracted considerable attention from the research community and industry over the last decade. Improvement of several Machine learning algorithms and classifier performances like weighted associative classifier were reported in the detection of cardiac abnormalities [7].

The elderly patients monitoring from indoor or outdoor locations had been presented by a real-time mobile healthcare system in [3]. A signal sensor and a smartphone were the primary components of the system. The bio-signal sensor data was transmitted to an intelligent server via GPRS/UMTS network for data collection. The system could perform in monitoring the mobility, vital signs, location, and condition of the elderly patient from a distant location. A fully functional wireless body area network (WBAN) system had been proposed in [8]. The designed system used medical bands to obtain physiological data from sensors. The author had chosen some medical bands in order to abate the interruption between the sensors and other existing devices. To increase the operating extent, the multi-hopping technique had been implemented and a medical gateway wireless board had been used in this regard.

Manpreet Singh *et al.* [9] designed a heart disease prediction system based on Structural Equation Modelling (SEM) and Fuzzy Cognitive Map (FCM). They validated the data of the Canadian Community Health Survey (CCHS) 2012 dataset. They used twenty significant attributes. To generate the weight matrix for

the FCM model, SEM was used which then predicted a possibility of cardiovascular diseases. An SEM model is defined with the correlation between CCC 121 along with 20 attributes; here CCC 121 is a variable which defines whether the respondent has heart disease. Prajakta Ghadge *et al.* [10] researched an intelligent heart attack prediction system using big data. Heart attack needs to be diagnosed timely and effectively because of its high prevalence. The main objective of this research article was to find a prototype of an intelligent heart attack prediction system that uses big data and data mining modeling techniques. This system could gather hidden knowledge concerning heart disease from a given heart disease database.

This work aims in developing a Decision Support System in heart disease detection that uses the data mining technique having best accuracy and performance among Naïve Bayes, Support Vector Machine, Simple Logistic Regression, Random Forest & Artificial Neural Network (ANN) etc. By using several cardiovascular system parameters such as age, blood pressure, ECG results, sex, and blood sugar, it is possible to measure the possibility of getting affected by heart disease [11]. For deriving the algorithm with the best accuracy in the detection and prediction of heart disease, a comparative analysis of chosen machine learning algorithms has been shown. This algorithm takes the medical parameters such as age, blood pressure, heartbeat, sex, ECG results, blood sugar etc. as input and shows the probability of getting affected by heart disease as output. This proposed system comprises the scheme and design of a web-based android application which uses an efficient machine learning technique to detect heart disease. It can serve as a very useful tool for doctors, patients, and medical students to diagnose heart disease. For diagnosis of fatal physiological conditions and symptoms such as heart attack requires 24 hours monitoring of patient's health after transferring from hospital to home. Using this application, the patient can input the current parameters of heart disease from anywhere on the application interface and view the risk level of getting heart disease. All of the parameters that are not real-time like blood sugar, Serum cholesterol, ECG results will be available in the doctor's prescribed report and there are some parameters like chest pain type and exercise-induced angina, which have to be self-measured periodically by the patient and input the values manually on the application interface. If any fatal situation is detected by the application, the patient can communicate with any doctor via video call and any registered doctor related to heart disease can be found by putting his phone number in the search bar.

In hospitals, there are provisions for continuous monitoring of critical care heart patients whereas after the release from the hospital patients normally go out of direct supervision. These patients need continuous monitoring of their health condition to reduce the risks of unwanted complications at least for a week or so. Hence, another objective of this work is the empirical design and implementation of the prototype of a continuous real-time cardiac health moni-

toring system by using Arduino based sensors, which can be developed commercially so that it can be attached to the patient's body. The sensor data can be transmitted to the server using the Wi-Fi module and saved in the server database. The sensor data can be updated every 10 seconds in the server and doctors with smartphones can view his patient's updated health status by using this application from anywhere. While any value of the heartbeat, temperature or humidity sensors goes beyond the threshold value, doctors will receive an alert message in the application as well as his phone instantly. Besides, family members and caretakers can also visualize the patient's real-time data through the application and if the patient remains in the ICU, the buzzer in the Arduino will alert them even though in case of network failure. When a patient will need consultancy or medication from a doctor from another distant location of the world, he/she can simply start live video streaming through the application.

3. Methodology and Data Analysis

In this study, an efficient machine learning algorithm was chosen from some available algorithms in a Java-based open access data mining platform (WEKA) to detect the presence or to decide the probability of having heart diseases from a large dataset. Then a continuous cardiac monitoring system design has been proposed using Arduino based microcontroller system. The step by step design approaches of the proposed system and the workflow of the complete system have been mentioned below:

- Collection and selection of different heart disease datasets in order to train various machine learning algorithms.
- Comparison of various data mining algorithm's accuracy and performance in predicting heart disease.
- Selection of the best algorithm from performance characteristics of the models to develop a cloud-based intelligent heart disease prediction mobile application.
- Storing of doctor and patient information following registration of patients and doctors separately through the application in a cloud-based server for analysis.
- A full-fledged tentative design of an android application with respective criteria has been shown in the Analysis and Results section.
- On the application interfaces, the patient can input all parameters of heart disease manually and they can input sensor data like heartbeat using a specific button. Hence, they can predict whether they have heart disease or not.
- After that, the patient can send the result to any doctor registered in the application in a report format by typing doctor's ID in the search option. Furthermore, He/she can start a live video call with the doctor if any critical situation will arise.
- A primary wireless patient health monitoring system is developed using Arduino with temperature, humidity, and heartbeat sensors in order to collect

real-time patient physiological data and to detect the critical situation of the patient.

- All the sensor real-time data are stored and updated on the cloud server database automatically after a specific time (every 10 sec) and if any unwanted value of the sensors is encountered then prescribed doctor from any location in the world is notified via mobile SMS using GSM module. In the case of ICU patients, an alarm system exists using a buzzer to alert the caretaker or nurses immediately.
- After receiving an alert message containing the current data of the sensors in his phone, the doctor will log in to the application with a view to checking the patient's previous physiological data and consulting appropriate medication via live video streaming.

4. Proposed System Architecture

This section depicts the overview of the proposed system and illustrates all of the components, techniques and tools are used for developing the entire system. To develop an intelligent and user-friendly heart disease prediction system, an efficient software tool is needed in order to train huge datasets and compare multiple machine learning algorithms. After choosing the robust algorithm with best accuracy and performance measures, it will be implemented on the development of the smartphone-based application for detecting and predicting heart disease risk level. Hardware components like Arduino/Raspberry Pi, different biomedical sensors, display monitor, buzzer etc. are needed to build the continuous patient monitoring system. **Figure 1** portrays the block diagram of the whole system workflow.

4.1. Hardware Components of the Patient Monitoring System

1) *Arduino/Genuino Uno*: Arduino/Genuino Uno is a microcontroller board where ATmega328P microcontroller is inserted. It has 14 digital input or output pins among them 6 can be used as PWM outputs and 6 analog inputs. Not only functional pins but also some other pins exist such as power pin, a 16 MHz quartz crystal oscillator, an ICSP header, a reset button, and a USB connection. Although there is a power jack, USB connection cable can also be used as a power supply. It has everything needed to provide support for the microcontroller, Simply it can be connected to a computer using USB cable or powered with an AC-to-DC adapter or battery.

2) *Heartbeat Sensor*: The heartbeat sensor has emerged on the postulation of light modulation on blood flow through the finger in each pulse. Any change of light intensity through that organ (a vascular region) is predicted with the rate of heart pulses and since light is also absorbed by blood, those signal pulses are equivalent to the heartbeat pulses [12]. It is constructed in such a way that gives a digital output of the heartbeat, while a finger is placed onto it. That digital output can also be connected to Arduino directly for measuring the Beats per

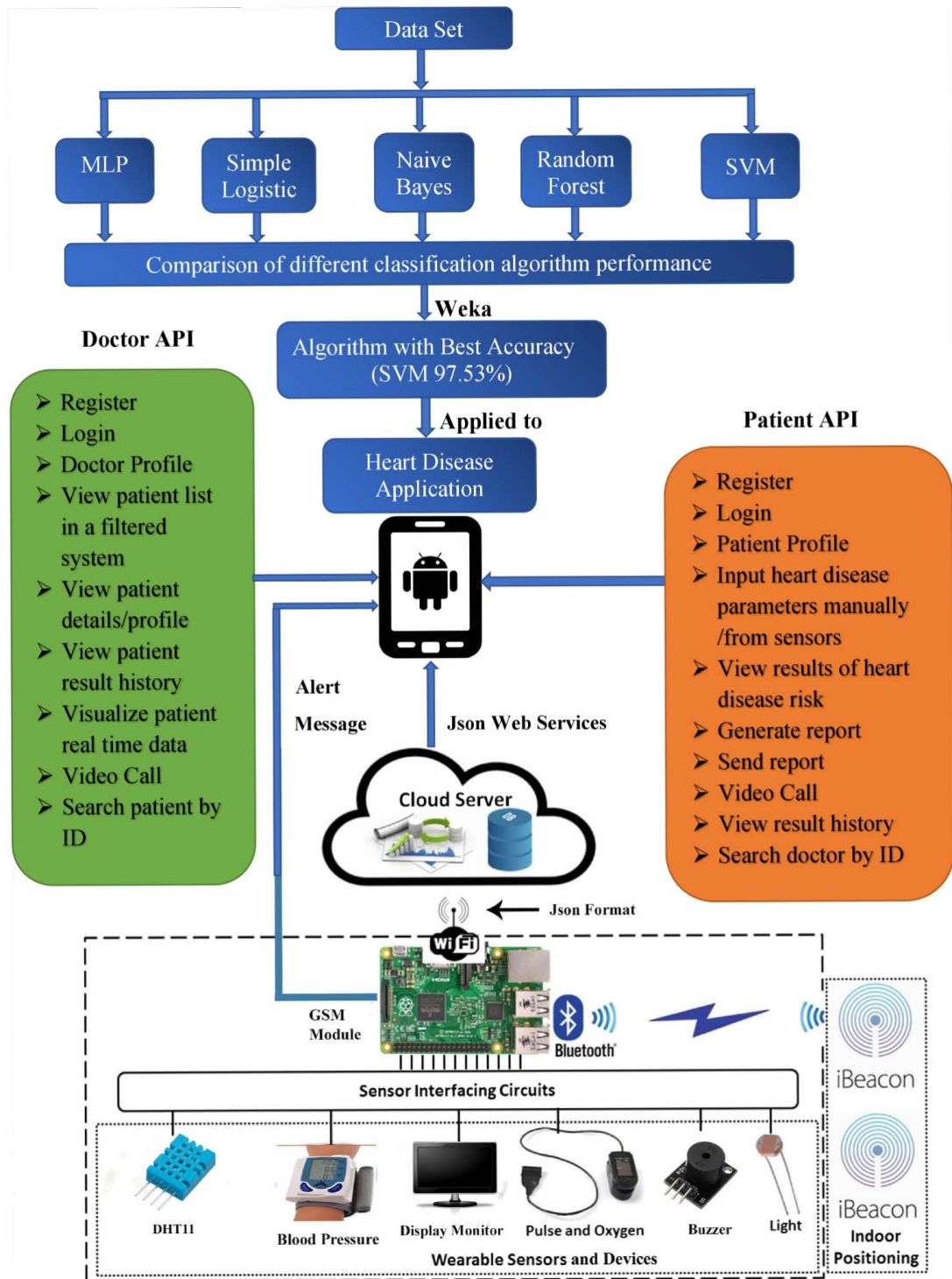


Figure 1. Overview of the proposed heart disease prediction and patient monitoring system.

Minute (BPM) rate.

3) *Electric Buzzer*: Buzzer is an electronic device used to generate alert sound. Here the buzzer is used to alert the caretaker during extreme condition. This sound indicates that the patient’s health is at immediate risk. Thus, nearest family members and caregivers are easily notified and help it to operate as a

real-time patient monitoring system.

4) *Temperature and Humidity Sensor (DHT11)*: DHT11 is a device which includes both humidity and temperature sensor. So, it can measure temperature and air humidity of a specific place whether at indoor or outdoor. Humidity and temperature are the important parameters to monitor patient's comfort or to check the patient's physical condition in that place as the patients feel uncomfortable due to rapid temperature and humidity changes which can also trigger abnormal cardiac responses [13].

4.2. Machine Learning Algorithms

Many data mining tools are provided to implement machine learning algorithms. Some frequently used, free and open source tools are WEKA, TANAGRA, RapidMiner, MATLAB, Apache Mahout etc. [14]. Among them, WEKA version 3.8.2 is used in this work. The full meaning of WEKA is Waikato Environment for Knowledge Learning. It's a computer program, which was developed at the University of Waikato in New Zealand to identify knowledge from raw data. It's an open source software that is issued under the General Public License. WEKA supports different types of data mining tasks, for example, classification, regression, feature selection, data pre-processing, clustering and visualization. WEKA was originally written in C, but later it was rewritten in Java and it's now compatible with the most computing platform. It is user-friendly and its graphical interface comprises quick set-up and operation.

More than twenty machine learning algorithms available in WEKA were applied to the UCI Machine Learning Repository Dataset and Statlog Database for Heart disease prediction [15] [16] to validate the model performance. Among those five with greater accuracy (*i.e.* more than 80%) are taken into consideration for performance measurement and they are described briefly below:

1) *Naive Bayes*: Naive Bayes is a surprisingly powerful algorithm for predictive modeling. It is a statistical classifier which assumes no dependency between attributes attempting to maximize the posterior probability in determining the class. Theoretically, this classifier has the minimum error rate, but may not be the case always. Inaccuracies are caused by assumptions due to class conditional independence and the lack of available probability data. This model is associated with two types of probabilities which can be calculated from the training dataset directly:

- a) The probability of every class.
- b) The conditional probability of each class with each x value.

According to Bayesian theorem $P(A|B) = P(A) * P(B|A) / P(B)$, where $P(B|A) = P(A \cap B) / P(A)$. Bayesian classifier calculates conditional probability of an instance belonging to each class, based on the above formula, and based on such conditional probability data, the instance is classified as the class with the highest conditional probability. If these probabilities are calculated, then the probabilistic model can be implemented to make predictions with new data us-

ing Naïve Bayes Theorem. When the data is real-valued it is likely to assume a Gaussian distribution (bell curve). Thus, these probabilities can easily be estimated. Naive Bayes is called naive because of assuming each input variable independent. This classifier algorithm uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of other attributes.

2) *Artificial Neural Networks*: Artificial neural networks also called Multilayer Perceptron are known as biologically inspired and it is capable of modeling extremely complex non-linear functions. ANNs are one of the major tools used in machine learning. As the name “neural” suggests, they are brain-oriented systems that are intended to duplicate the way how humans learn. Neural networks consist of 3 layers of input, output and hidden layer. In most cases, a hidden layer comprises units that transform the input to a pattern that the output layer manipulates. ANN’s are excellent tools for the purpose of finding patterns that are so complex or ambiguous to a human programmer to extract and teach the machine how to recognize. Neural networks are in use since the 1940s and over the last decades they have become an important part of artificial intelligence because of the arrival of a new technique, which is called “backpropagation”, that allows networks know how to adjust their hidden layers of neurons in cases where the outcomes don’t match with the creator’s expectation. In **Figure 2** the interconnection between the layers is shown.

3) *Support Vector Machine*: SVM is a technique for ramification of both linear and non-linear data. It applies a non-linear mapping method so that it can transform the training data into a higher dimension. A hyperplane is a kind of line which separates the input variable space in SVM. The hyperplane can separate the points in the input variable space containing their class that is either 0 or

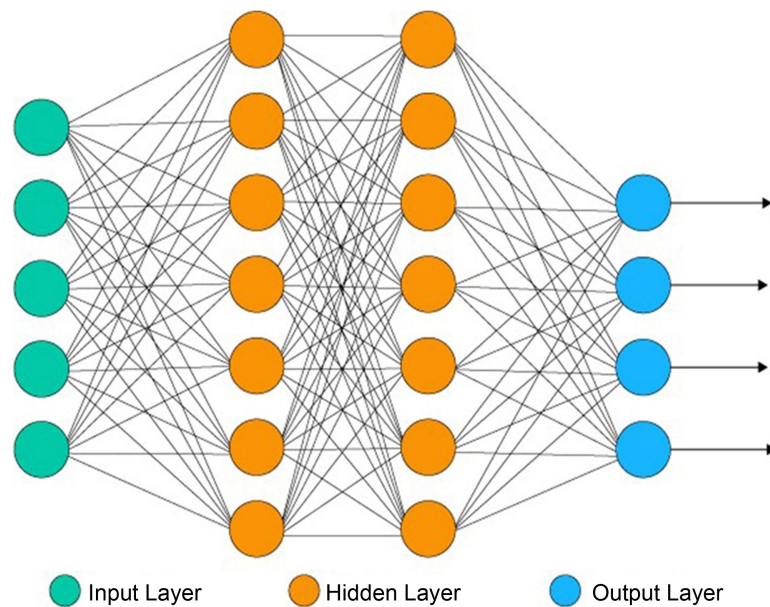


Figure 2. Diagram of an artificial neural network.

1. In two-dimensions, one can visualize this as a line and it is assumed that each input points can be completely separated by this line. The distance between the hyperplane and adjacent data coordinates is called margin. The line which has the largest margin can distinguish between the two classes is known as the optimal hyperplane. These points are called support vectors, as they define or support the hyperplane. In practice, there is an optimization algorithm which is used to calculate the values for the parameters that maximize the margin. **Figure 3** depicts the feature transformation process.

4) *Random Forest*: Random Forest is one of the most renowned and most powerful machine learning algorithms. It is one kind of machine learning algorithm that is called Bagging or Bootstrap Aggregation. In order to estimate a value from a data sample such as mean, the bootstrap is a very powerful statistical approach. Here, lots of samples of data are taken, the mean is calculated, after that all of the mean values are averaged to give a better prediction of the real mean value. In bagging, the same method is used, but instead of estimating the mean of every data sample, decision trees are generally used. Here, numerous samples of the training data are considered and models are generated for every data sample. While a prediction for any data is needed, each model gives a prediction and these predictions are then averaged to get a better estimation of the real output value.

5) *Simple Logistic Regression*: Logistic regression is a technique of machine learning which is taken from the field of statistics. This method can be used for binary classification where values are distinguished with two classes. Logistic regression is similar to linear regression where the goal is to calculate the values of the coefficients within every input variable. Unlike linear regression, here the prediction of the output is constructed using a non-linear function which is called a logistic function. The logistic function transforms any value within the range of 0 to 1. The predictions made by logistic regression are used as the probability of a data instance concerning to either class 0 or class 1. This can be necessary for problems where more rationale for a prediction is needed. Logistic regression works better when attributes are unrelated to output variable and

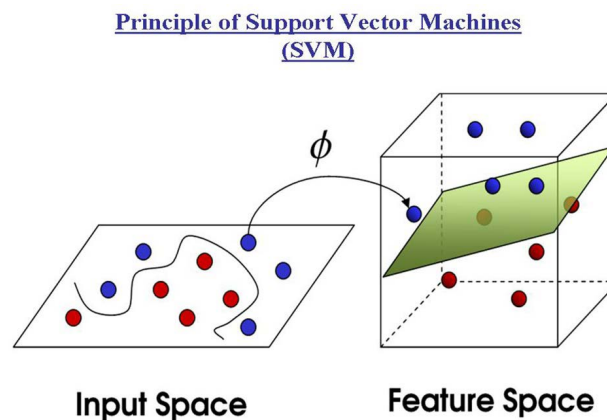


Figure 3. Principles of support vector machine operation.

attributes correlated to one another are removed.

4.3. Cloud Server

In cloud server, there is a core, a large database that has enough space in order to accommodate huge amounts of data from the different sensors so that the history of the system user can easily be tracked. A wide range of data analyzing algorithms is interfaced with the database and APIs such as Google Sheets for data visualization. A data flow diagram of the cloud server is presented in **Figure 4**.

4.4. Data Source

Two publicly available heart disease databases having the same type and number of attributes are merged to form a bigger dataset and avail higher accuracy. They are the Cleveland Heart Disease dataset [15] consists of 303 records and Statlog Heart Disease dataset consists of 270 records [16] with 13 similar features as shown in **Table 1**. By similar features it was meant that both the Cleveland Heart Disease dataset and Statlog Heart Disease dataset have these features used for heart disease detection and prediction. Both the datasets were merged to increase the robustness of the designed model. After merging these two datasets, instances containing missing values are removed. After that, 566 instances are left and used for model validation. The data mining tool WEKA version 3.8.2 has been used for pre-processing and classification using the above-mentioned Machine learning algorithms.

4.5. Attribute Documentation

The dataset taken for data mining application includes 13 kinds of input

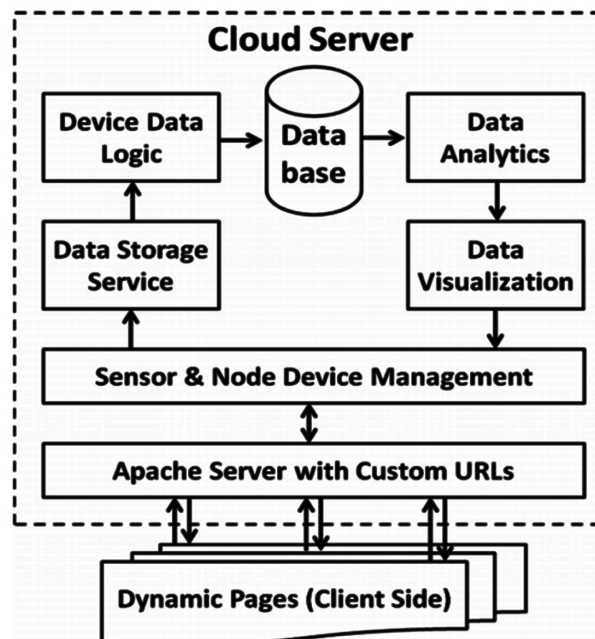


Figure 4. Data flow diagram of cloud server.

Table 1. Explanation of 13 Input Attributes used for model formation and validation.

No.	Attribute	Description
1	Age	Age in years
2	Sex	Male = 1, female = 0
3	Cp	Chest pain type (typical angina = 1, atypical angina = 2, non-anginal pain = 3, asymptomatic = 4)
4	Trestbps	Resting blood sugar (in mm Hg in case of admission to hospital)
5	Chol	Serum cholesterol in mg/dl
6	Fbs	Fasting blood sugar > 120 mg/dl (true = 1, false = 0)
7	Restecg	Resting electrocardiographic results (normal = 0, having ST-T wave abnormality = 1, left ventricular hypertrophy = 2)
8	Thalach	Maximum heart rate
9	Exang	Exercise-induced angina
10	Old peak	ST depression induced by exercise comparative to rest
11	Slope	Slope of the peak exercise ST segment (upsloping = 1, flat = 2, down sloping = 3)
12	Ca	Number of major vessels which are colored by fluoroscopy
13	Thal	Normal = 0, fixed defect = 2, reversible defect = 3

attributes which are related to different cardiovascular system parameters. For the design of Cloud-based classification and prediction system a Key attribute was designed (*i.e.* patient's mobile no which acts as the patient's unique identifier). Predictable attributes are given a numerical scale for determining the two classified groups (*i.e.* Healthy and Having heart disease). **Table 1** shows the input attributes containing a total of 13 physiological features taken from the two previously mentioned datasets used for any type of cardiac abnormality prediction.

1) *Input Attribute*

2) *Key Attribute for Cloud-Based System*

Mobile No: Patient's Identification Number.

A mobile number is used as username during doctor/patient registration process through the application. This number can uniquely identify any patient or doctor. While using the app by typing respective mobile number on the search bar, any doctor or patient ID can be found and communicated easily.

3) *Predictable Attribute*

Num: Value = 0 (No Heart Disease).

Value = 1 - 4 (Has Heart Disease).

All machine learning algorithms predict this attribute from input attributes on the classification process. This attribute values are divided into two groups, the value 0 expresses that the patient has no heart disease and values 1 - 4 are considered as value 1 which indicates that the patient has heart disease. Then the result analysis and performance measure processes are done by classifying the da-

taset records into these two groups.

5. Result Analysis

This section discusses the findings of the proposed technique for choosing the best machine learning algorithm and the operation of the cardiac monitoring system.

5.1. Performance Analysis of Machine Learning Algorithms

There are various kinds of algorithms are available, which can be applied to the dataset. More than 10 algorithms are applied to the dataset, among those only 5 important algorithms are analyzed and discussed which showed an accuracy level of more than 80%. The algorithms are applied to the data set by using 10-fold Cross-Validation on WEKA version 3.8.2 for measuring the performance of the machine learning algorithms. The following measures are calculated for performance analysis:

$$\text{Precision} = \frac{tp}{tp + fp} \quad (1)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (2)$$

$$F_{\text{score}} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

$$\text{Accuracy} = \frac{tp + tn}{n} \quad (4)$$

$$\text{Sensitivity} = \frac{tp}{tp + fn} \quad (5)$$

$$\text{Specificity} = \frac{tn}{tn + fp} \quad (6)$$

Here,

n = Total number of instances, tp = true positive, tn = true negative,
 fp = false positive, fn = false negative.

All of the performance measures of the algorithms are listed in **Table 2**. **Figure 5** shows a sample WEKA output screen indicating all the model parameters for reference.

5.2. Cloud-Based App Design

A cloud-based system was proposed where the patients can upload the physiological data for checking the status of their cardiac health. **Figure 6** shows the API of “Heart Doctor” mobile application indicating how patients and doctors initially will register on the heart disease application. In **Figure 7** the layout will appear after patient log in and then he/she can input individual heart disease parameters and observe the results of heart disease diagnosis.

Table 2. Performance analysis of five machine learning algorithms (in Percentage).

Evaluation Criteria	Naive Bayes	SVM	Random Forest	Simple Logistic	ANN
Accuracy	86.40	97.53	95.76	95.05	77.39
True Positive Rate/Recall/Sensitivity	86.40	97.50	95.80	95.10	77.40
True Negative Rate/Specificity	82.88	94.94	92.61	93.77	63.42
False Positive Rate/False Out	14.20	2.90	4.80	5.20	25.00
False Negative Rate/Miss Rate	10.68	0.32	1.62	3.88	11.00
Precision	86.40	95.95	95.90	95.10	78.30
F-score	86.40	96.72	95.70	95.00	76.90

```

Classifier output
Time taken to build model: 0.53 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      552      97.5265 %
Incorrectly Classified Instances    14      2.4735 %
Kappa statistic                    0.9499
Mean absolute error                 0.0247
Root mean squared error             0.1573
Relative absolute error             4.9889 %
Root relative squared error         31.5879 %
Total Number of Instances          566

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.997   0.051   0.960     0.997   0.978     0.951   0.973    0.958     0
                0.949   0.003   0.996     0.949   0.972     0.951   0.973    0.969     1
Weighted Avg.   0.975   0.029   0.976     0.975   0.975     0.951   0.973    0.963

=== Confusion Matrix ===

  a  b  <-- classified as
308  1  |  a = 0
 13 244 |  b = 1

```

Figure 5. Weka display screen showing different parameters for SVM.

Figure 8 shows the online connectivity of the patients to check their status. After logging into the application, the layout in **Figure 8** will appear on the doctor profile. Here, he/she can view the patient list with priority system according to their risk level and he can also view the patient's details and previous diagnosis results. Besides, after clicking on the monitoring system button he/she can visualize each patient's health status and real-time physiological data.

5.3. Experimental Setup of the Proposed Cardiac Monitoring System

A heartbeat sensor MOD-00158 module is used and connected to Arduino Uno via cables to monitor real-time heartbeat (BPM). It can be worn on the finger or

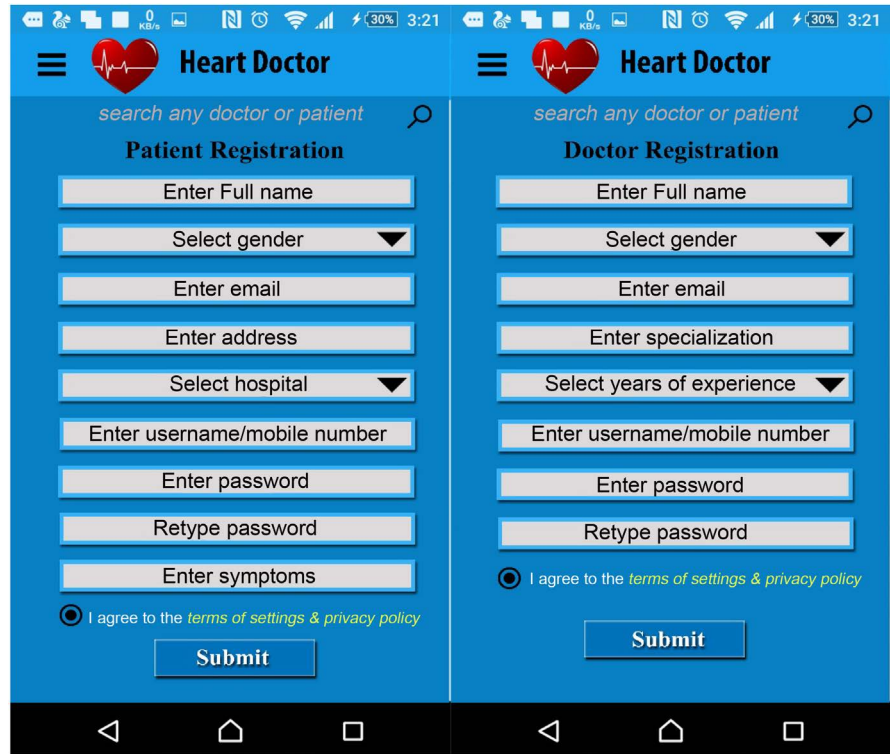


Figure 6. Patient and doctor registration process on the application.

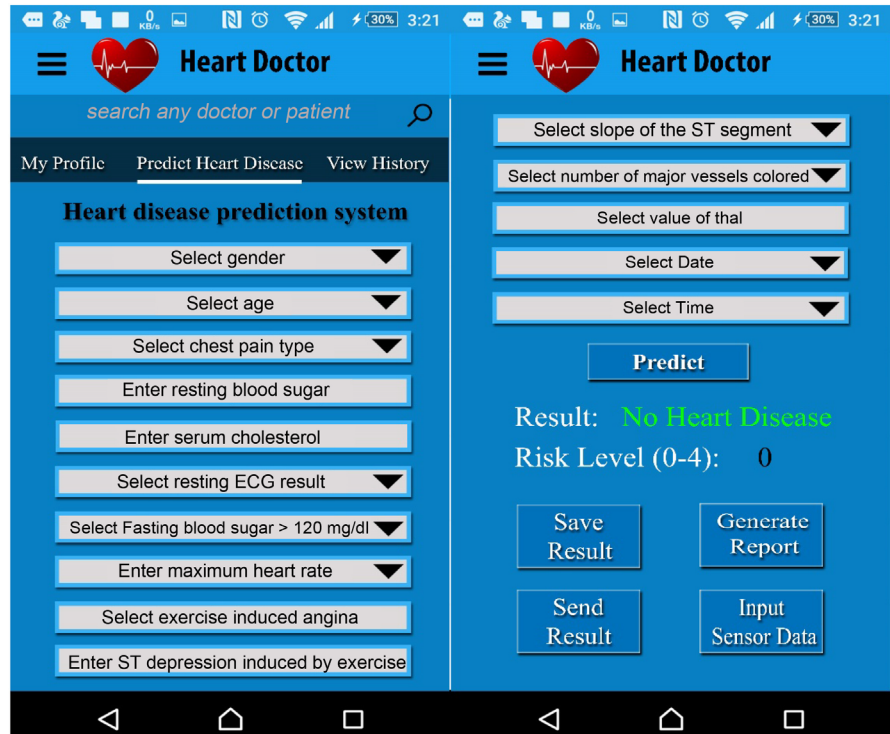


Figure 7. Patient data I/O and operation on result.

earlobe to measure the real-time heartbeat. On the other hand, a combination of humidity and temperature sensor module DHT11 is also connected to the Arduino

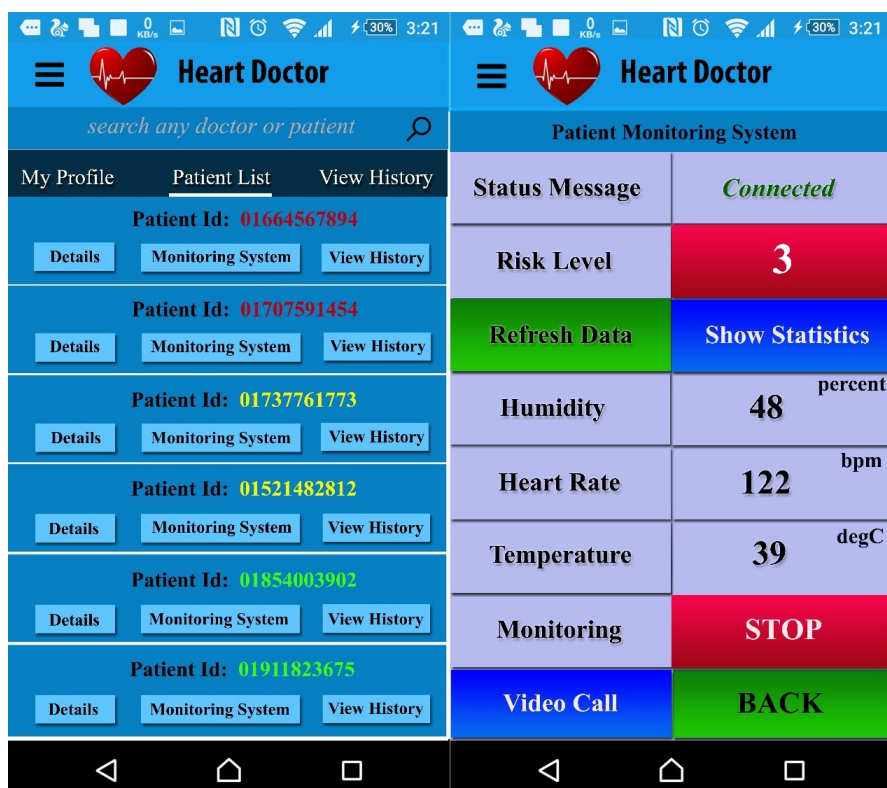


Figure 8. Patient list according to priority and live monitoring API.

Uno board via some jumper wires. The humidity and temperature sensors are essential for measuring the patient's body temperature and humidity and when the patient passes critical condition. In that condition, the patient can't be able to share his/her feelings with other. In that time, this system will read the air components such as humidity and temperature and also it will give a comment about the patient's feelings in that condition. A buzzer is connected in this system to give an alarm when heartbeat, temperature, humidity or any of these parameters crosses the normal threshold value, it will alert the patient's nearest caregiver. It can also send an alert message to the doctor via a wireless module and a GSM module, and then the doctor may also observe the patient's condition through live video streaming by using the application. Finally, a serial monitor is also connected with this system to observe those parameters live. In **Figure 9** the full setup of the hardware components has been shown.

Figure 10 shows the display of the recordings of the patient's data from the Arduino based continuous monitoring system. **Table 3** indicates the levels of temperature and moisture along with heart rate to generate a decision-making system for the Continuous Patient Monitoring system. This system is quite effective for monitoring heart disease patients who were released from the hospital but need continuous supervisions for his/her safety.

6. Discussion and Conclusion

From **Table 2**, it is clear that SVM, Random Forest and Simple Logistic models

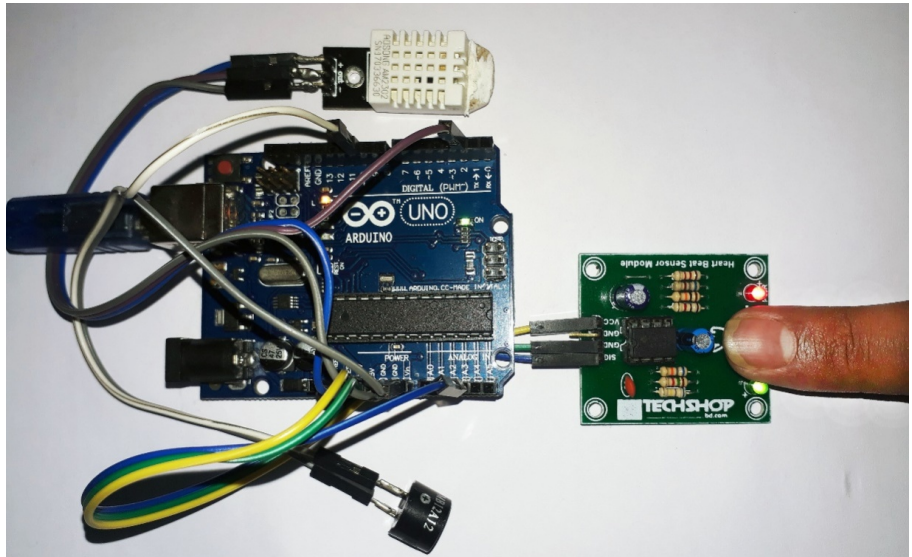


Figure 9. Arduino Uno with the necessary sensors.

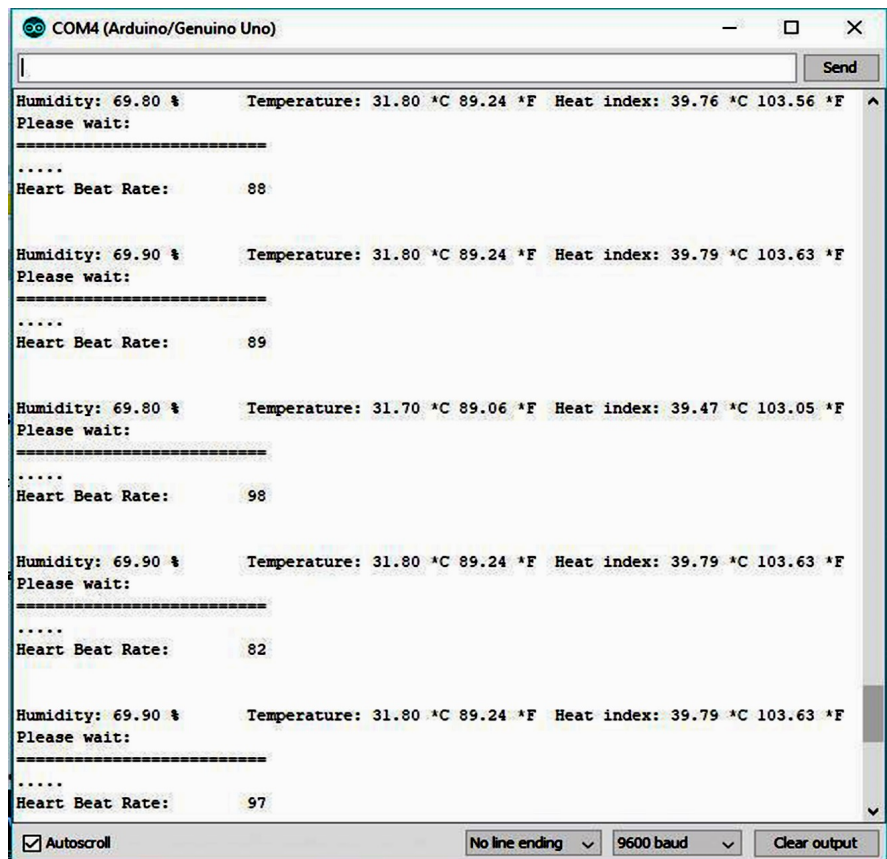


Figure 10. Real-time sensor data on a serial monitor in Arduino-based system.

showed higher accuracy rates of more than 95 percent which make them considerable models for biomedical applications of disease detection and prediction. Moreover, to determine the best performing model for the considered databases in this study, other analysis criteria are considered. The accuracy, sensitivity,

specificity, precision, and F-score of SVM is higher than both Random Forest and Simple Logistic models. Besides, the miss rate of SVM is lowest. The number of features also affects the performance of the model classification and prediction capabilities. The unpublished results of this work indicated that when features numbers were reduced, the model performance measures were also reduced with SVM still showing the best performance. The same analysis was performed in Python platform where SVM also showed the best performance among the models with Radial Basis Kernel Function. Therefore, it is decisive that Support Vector Machine is the most efficient algorithm to be implemented on the heart disease prediction system as found in our study when the 13 features were considered. In the previous research works [2]-[14] [17] [18] [19] concerning comparison among various machine learning algorithms, it was found that no algorithm attained an accuracy level of more than 90 percent in heart disease prediction using the same number and types features as used in this study. Moreover, in this study, two datasets consisting of the same number and type of attributes were merged. Therefore, the chosen model will be more robust than that of presented in other studies. **Table 3** explains the relationship among the temperature, humidity, heart rate and different decisions of action pertaining to risk level and human comfort that can be applied to the designed Continuous Patient Monitoring system. This design validates the possibility of an integrated Cardiac patient monitoring system which can be used by the patient at home environment and enables the patient with an online centralized monitoring system where the doctors/nurses can provide real-time suggestions.

In future works, Photoplethysmography (PPG) based blood pressure sensor module or electronic sphygmomanometer can also be connected to the Arduino which will be capable of transmitting real-time data to the server. This sensor is not added to the designed patient monitoring system due to the unavailability of a clinically recognized system at this moment, although huge research is going on the development of PPG based blood pressure monitor [20]. Though a model of the cloud-based heart disease detection application is represented here, the future works will be focused on the development of a dedicated server and data

Table 3. Boolean table for decision making about the condition of patient's surroundings.

Temp. sensor	Humidity	Human Perception	Pulse rate sensor (per min.)	Action taken	Risk level
<37°C	31% - 41%	Comfortable	60 to 100	No action	0/1
37°C - 38°C	41% - 46%	Comfortable for most people	40 - 60 or 100 - 120	Inform family	2
>38°C	46% - 52%	Somewhat uncomfortable for most people at upper age	40 - 60 or 100 - 120	Inform local doctor	3
>38°C	>52%	Very humid, Extremely uncomfortable	<40 or >120	Inform emergency	4

base for this type of patient monitoring application. If the proposed application can be integrated and developed successfully, it will be available in the Android play store consequently. Thus, any patient or doctor from any region of the globe will be capable of installing this application and use this application for heart disease prediction as reported in [21]. Alongside heart disease, this system can also be used for any disease patient monitoring purposes.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Hazra, A., Mandal, S., Gupta, A. and Mukherjee, A. (2017) Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. *Advances in Computational Sciences and Technology*, **10**, 2137-2159.
- [2] Patel, J., Upadhyay, P. and Patel, D. (2016) Heart Disease Prediction Using Machine learning and Data Mining Technique. *Journals of Computer Science & Electronics*, **7**, 129-137.
- [3] Chavan Patil, A.B. and Sonawane, P. (2017) To Predict Heart Disease Risk and Medications Using Data Mining Techniques with an IoT Based Monitoring System for Post-Operative Heart Disease Patients. *International Journal on Emerging Trends in Technology (IJETT)*, **4**, 8274-8281.
- [4] Weng, S.F., Reys, J., Kai, J., Garibaldi, J.M. and Qureshi, N. (2017) Can Machine-Learning Improve Cardiovascular Risk Prediction Using Routine Clinical Data? *PLoS ONE*, **12**, e0174944. <https://doi.org/10.1371/journal.pone.0174944>
- [5] Zhao, W., Wang, C. and Nakahira, Y. (2011) Medical Application on Internet of Things. *IET International Conference on Communication Technology and Application (ICCTA 2011)*, Beijing, 14-16 October 2011, 660-665.
- [6] Chiuchisan, I. and Geman, O. (2014) An Approach of a Decision Support and Home Monitoring System for Patients with Neurological Disorders Using Internet of Things Concepts. *WSEAS Transactions on Systems*, **13**, 460-469.
- [7] Soni, J., Ansari, U. and Sharma, D. (2011) Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers. *International Journal on Computer Science and Engineering (IJCSE)*, **3**, 2385-2392.
- [8] Yuce, M.R. (2010) Implementation of Wireless Body Area Networks for Healthcare Systems. *Sensor and Actuators A: Physical*, **162**, 116-129. <https://doi.org/10.1016/j.sna.2010.06.004>
- [9] Singh, M., Martins, L.M., Joanis, P. and Mago, V.K. (2016) Building a Cardiovascular Disease Predictive Model Using Structural Equation Model and Fuzzy Cognitive Map. *IEEE International Conference on Fuzzy Systems (FUZZ)*, Vancouver, 24-29 July 2016, 1377-1382. <https://doi.org/10.1109/FUZZ-IEEE.2016.7737850>
- [10] Ghadge, P., Girme, V., Kokane, K. and Deshmukh, P. (2016) Intelligent Heart Attack Prediction System Using Big Data. *International Journal of Recent Research in Mathematics Computer Science and Information Technology*, **2**, 73-77.
- [11] Shouman, M., Turner, T. and Stocker, R. (2012) Using Data Mining Techniques in Heart Disease Diagnosis and Treatment. *Electronics, Communications, and Computers*, Alexandria, 173-177.

-
- [12] Alemdar, H. and Ersoy, C. (2010) Wireless Sensor Networks for Healthcare: A Survey. *Computer Networks*, **54**, 2688-2710.
<https://doi.org/10.1016/j.comnet.2010.05.003>
- [13] Antony, G. and Francis, S. (2015) Patient Monitoring System Using Raspberry PI. *International Journal of Science and Research*, **6**, 687-689.
- [14] Kumari, M. and Godara, S. (2011) Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction. *International Journal of Computer Science and Technology*, **2**, 304-308.
- [15] UCI Machine Learning Repository: Heart Disease Data Set.
<http://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- [16] Statlog Database.
<https://data.world/uci/statlog-heart/workspace/file?filename=heart.dat.txt>
- [17] Soni, J., Ansari, U., Sharma, D. and Soni, S. (2011) Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications*, **17**, 43-48.
- [18] Dangare, C.S. and Apte, S.S. (2012) Improved Study of Heart Disease Prediction System Using Data Mining Classification Techniques. *International Journal of Computer Applications*, **47**, 44-48.
- [19] Masethe, H. and Masethe, M. (2014) Prediction of Heart Disease Using Classification Algorithms. *Proceedings of the World Congress on Engineering and Computer Science*, San Francisco, 809-812.
- [20] Elgendi, M., Liang, Y. and Ward, R. (2018) Toward Generating More Diagnostic Features from Photoplethysmogram Waveforms. *Diseases*, **6**, 20.
<https://doi.org/10.3390/diseases6010020>
- [21] Ahmed, F. (2017) An Internet of Things (IoT) Application for Predicting the Quantity of Future Heart Attack Patients. *International Journal of Computer Applications*, **164**, 36-40.