International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume:10 Issue:6 | June 2023 www.irjet.net p-ISSN: 2395-0072

Angina Pectoris Predicting System using Embedded Systems and Big Data

Dangeti Sai Chaithrik, Chiluka Harshith Reddy

Dangeti Sai Chaithrik, Dept. of Electronics and communication Engineering, VIT university, Tamil Nadu, India
Chiluka Harshith Reddy Dept. of Electronics and communication Engineering, VIT university, Tamil Nadu, India

---***---

Abstract - *The proposed system paves a way to predict the disease with the seamless integration of both hardware and Machine learning. The embedded System consists of components like heart rate sensor, SpO2 sensor, ESP8266 WIFI-module and ECG sensor for collecting the patient data modules that includes patient heart rate, rest ECG, temperature and SpO2 and are visualized using Thing Speak cloud. The heart rate ranges from 20BPM to 150BPM where BPM above 82 is considered as high or abnormal heart rate and BPM below 60 is considered as low heart rate. After collection this data is fed into the five machine learning algorithms which includes 1. logistic regression 2. Random Forest 3. K-NN 4. Linear SVM 5. Naive bayes for training and testing the data. The efficiency is highest for K-NN with 0.8703 and Random Forest with 0.9640 test accuracy respectively so the proposed system uses these two models for the prediction. In the web page both K-NN and Random Forest are incorporated to predict if the patient is suffering from disease or not from various field ID's which are inputted by the end user.*

***Key Words*: Anaconda, BeatsPerMinute (BPM), Classifier, Electrocardiography (ECG), prediction rate, Thingspeak.**

1.INTRODUCTION

IoT refers to the network of physical devices, sensors, and connectivity that enable these objects to connect and exchange data. In the healthcare industry, IoT devices can be used to monitor patients in real-time, collect health data, and provide remote care. ML, on the other hand, that enables machines to learn from data and make predictions based on that data. In the healthcare industry, ML can be used to analyse vast amounts of patient data to identify patterns, predict outcomes, and make personalized treatment recommendations.

The combination of IoT and ML has the potential to revolutionize the healthcare industry and improve the lives of patients and healthcare providers. The literature reported in [1-17] deals with various ways to collect various health parameters like heart rate, NodeMCU, ECG of the patients and the application of various machine learning models in training and testing the acquired data. Health monitoring using sensors and predicting framework using IOT portrays the intercorrelation and assortment of the patient health information [1]. Establishing IOT environment with wearable sensor devices like Heart rate senor, ECG for the data collection and the collected data is sent to the Arduino for processing of data [2]. The processed data is sent into the cloud like Thingspeak for continuous visualization and to get a true picture of the patient's heart rate and temperature variations however this paper does not propose a way to integrate the ECG graph into the Thingspeak [3]. The pulse rate is divided into three categories namely low heart rate where BPM<60, Normal heart rate where BPM lies between 60 to 100 and high pulse rate where BPM>100 these variations can better help to visualize the data and these classifications can be done in the Arduino initialization [4]. It is crucial to create portable heart detection systems that can process signals in real time and analyse an ECG in order to keep track of high-risk patients and help doctors make judgements [5]. Unlike the unsupervised learning the supervised learning the algorithm is trained on labelled data that includes both input and output features and in turn give better test accuracy [6]. logistic regression predicts heart disease by calculating the likelihood of the outcome variable (heart disease) based on a collection of predictor factors using a logistic function (sigmoid curve) [7]. The Random Forest technique builds many decision trees, each of which is trained using a different random subset of the data. Each tree in the forest predicts whether a certain patient has heart disease or not, and the combined projections of all the trees in the forest yield the final result [8]. The K-NN algorithm looks for the "k" closest patients in the training data based on the similarity of their features to the new patient's features and the "k" closest patients are then used to make a prediction about whether the new patient is likely to have heart disease or chest pain [9]. The linear SVM algorithm works by finding the best hyperplane that separates the data into different classes. In the case of heart disease prediction, the algorithm tries to find the best hyperplane that separates the patients who have heart disease from those who don't [10]. Naive Bayes works by using Bayes' theorem, which is a statistical formula that calculates the probability of an event occurring based on prior knowledge of conditions that might be related to the event. In the case of heart disease prediction, the algorithm tries to calculate the probability that a patient has heart disease [11]. K-NN is a non-parametric algorithm, which means that it does not make any assumptions about the underlying distribution of the data. This allows the algorithm to be highly flexible and adapt well to different types of data [12]. To lessen overfitting and increase algorithmic accuracy, each decision tree in the Random Forest is trained using a random selection of features and a subset of the data [13]. After training and testing the data, the best algorithms are selected and the model is saved using python pickling [14].

1.1 Technical Specifications

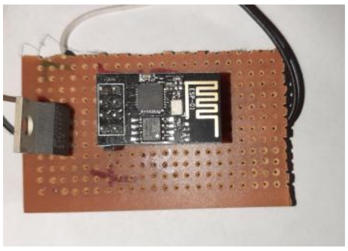
The Angina Pectoris Prediction System gathers metrics on the patient's heart rate, blood pressure, and other vitals using IoT sensors like heart rate sensors and ECG sensors connected to Arduino Uno and delivers the information over wifi to the Thingspeak api platform. Five machine learning algorithms are used to process and analyze the acquired data. Pre-trained machine learning models are stored on the website as pickle files and are used to make predictions. The system is intended to be precise and effective, giving medical practitioners access to results right away.

Hardware specifications:

Arduino Uno: The Arduino UNO is a microcontroller board based on the ATmega328P. It contains 6 analogue inputs, a 16 MHz ceramic resonator, 14 digital input/output pins (six of which may be used as PWM outputs), a USB port, a power connector, an ICSP header, and a reset button. It comes with everything needed to support the microcontroller; to get started, just plug in a USB cable, an AC-to-DC converter, or a battery.

Heart rate Sensor: A heart-rate sensor for Arduino, the Pulse Sensor, is plug-and-play. Students, artists, athletes, makers, game developers, and mobile app developers may all use it to quickly incorporate real-time heart rate data into their works. An integrated optical amplifier and noise-reduction circuit sensor is the key component. Attach the Pulse Sensor to your fingertip or earlobe. Then you may read heart rate by plugging it into your Arduino.

ESP8266 WiFi module: ESP8266 is a system on chip (SoC) module that supports Wi-Fi. It makes embedded applications internet-connected. For connection with the server/client, it interacts using the TCP/UDP protocol. It capable of Analog-to-digital conversion (10-bit ADC), Serial Peripheral Interface (SPI) serial communication protocol, 2.4 GHz Wi-Fi (802.11 b/g/n, supporting WPA/WPA2), general-purpose input/output (16 GPIO), Inter-Integrated Circuit (I2C) serial communication protocol, I2S (Inter-IC Sound) interfaces with DMA (Direct Memory Access) (sharing pins with GPIO), and UART (on dedicated pins, plus a transmit-only UART can be enabled on GPIO2).



LCD Display:

A sort of display that makes use of liquid crystals is the LCD (Liquid Crystal Display). There is a 16-pin interface on the LCD panel. The following is a discussion of the 16 pins on the LCD display:

1. The memory of the LCD that we send the data to is controlled by the Register Select (RS) pin. Either the data register or the instruction register is options. The future instruction, which is stored in the instruction register, is sought after by the LCD.

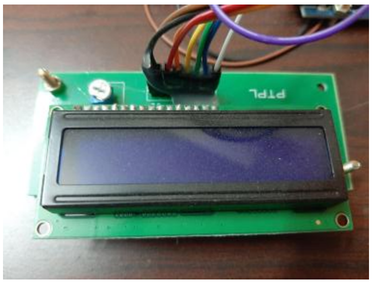
2. The reading or writing mode is chosen by the Read/Write pin.

3. The writing to the registers can be enabled by using the Enable (E) mode. When the mode is HIGH, it transmits the data to the data pins.

4. D0 to D7 these eight data pins have the following designations: D0, D1, D3, D3, D4, D5, D6, and D7. We have the option of setting the data pin's status to HIGH or LOW.

5. The LCD's Ground pin is pin 1, and the Vcc, or voltage supply, pin, is pin 2. The VEE, or contrast pin, is located on LCD pin 3.

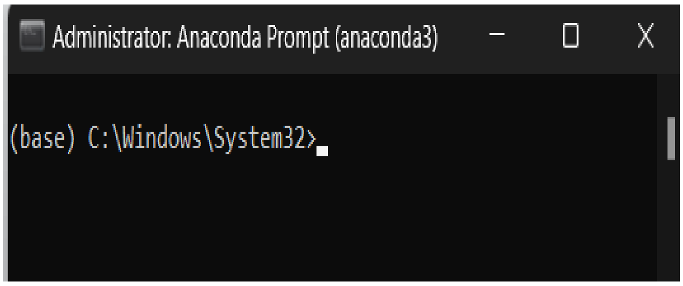
6. Backlight pins (Bklt+ and Bklt-) are another name for the A and K pins.



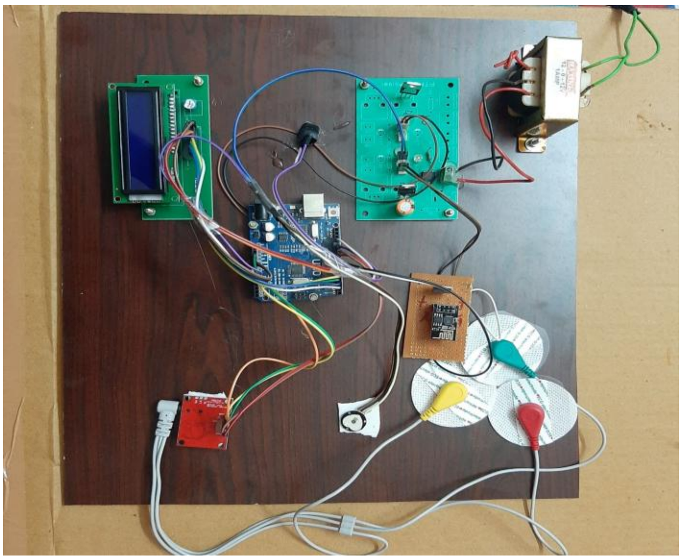
1.2 Software Specifications

1. Visual Studio: Microsoft's Visual Studio is an integrated development environment (IDE). Software applications, such as websites, web applications, online services, and mobile applications, are developed using it. Microsoft's software development platforms, including Windows Store, Windows Presentation Foundation, Windows API, and Windows Forms, are used by Visual Studio. Both native and managed code can be generated by it.

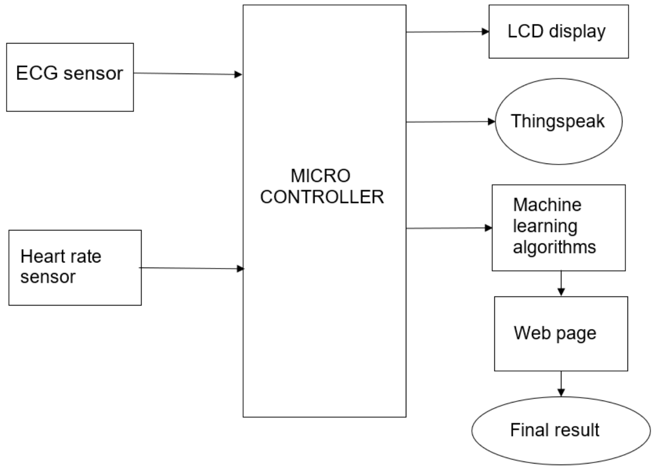
2. **Anaconda prompt:** With more than 720 open-source programmes, Anaconda Distribution is a free, simple-to-install package management, environment manager, and Python distribution. The command line interface for Anaconda Distribution is called Anaconda Prompt. Linux and macOS both have a command line interface called Terminal.



Working Prototype:



Block Diagram:



Description:

The Ecg sensor has been connected to the digitalRead pins of 10 and 11 for the LO+ and LO- detections. The heart rate sensor has been connected to the analog pin 0. The serial communication baud rate is fixed to be 9600. The beats per minute has been set in between 20 to 150 in the myBPM function call where the readings are recorded. The output is displayed in the LCD connected to the pins 7, 6, 5, 4, 3, 2, 1. The pins 8 and 9 are given to the receiver and transmitter module. The IP address of the thingspeak cloud is set to "184.106.153.149". The changes in the data can be viewed in the thingspeak open source with a delay of 500ms.

LCD Interfacing:

The initialization of the LCD has been given a delay of 1000ms and the variations in the heart rate values has been given a delay of 500ms.

Table -1: Pin description for LCD

Pin	Symbol	I/O	Description
1	Vss	–	Ground
2	Vcc	--	+5V power supply
3	VEE	–	Power supply to control contrast
4	RS	I	RS=0 to select command register RS=1 to select data register
5	R/W	I	R/W=0 for write R/W=1 for read
6	E	I/O	Enable

7	DB0	I/O	The 8-bit data bus
8	DB1	I/O	The 8-bit data bus
9	DB2	I/O	The 8-bit data bus
10	DB3	I/O	The 8-bit data bus
11	DB4	I/O	The 8-bit data bus
12	DB5	I/O	The 8-bit data bus
13	DB6	I/O	The 8-bit data bus
14	DB7	I/O	The 8-bit data bus

Design Flow: Design flow consists of four parts Data-collection, Data-preprocessing, Efficient model, prediction part.

Table-2: Design flow

Data-Collection (1)	Data-Preprocessing (2)
Efficient model (3)	Prediction (4)

Steps Involved in the working prototype:

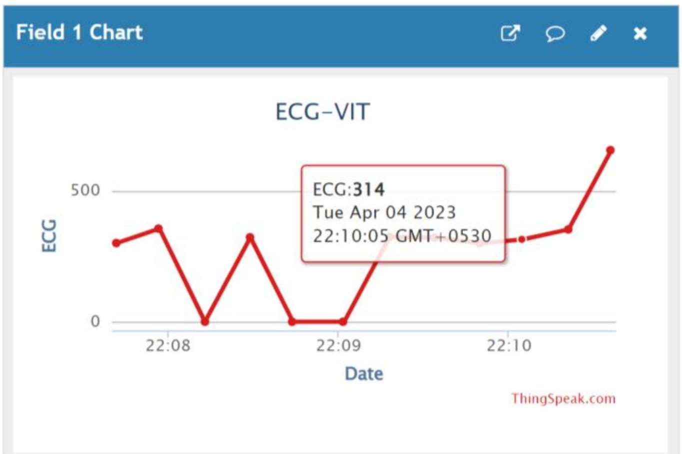
Step-1: Initialization

Step-2: Reading the BPM with a delay of 500ms

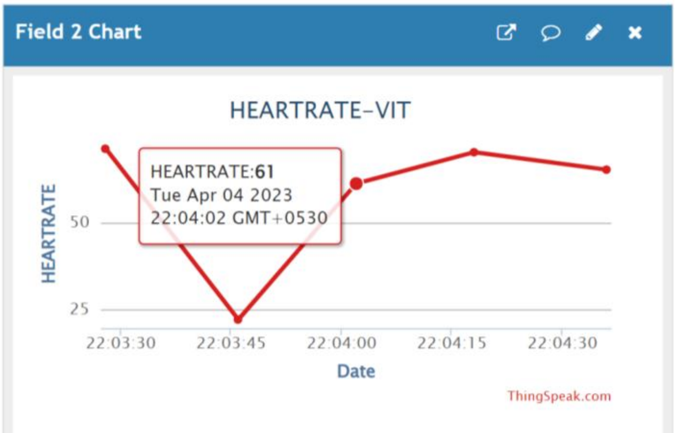
Step-3: Reading the ECG values with a delay of 500ms

Step-4: Visualizing and downloading the field id's in the CSV format

ECG graph:



Heart Rate:



A varied dataset can help ensure that the model is not biased towards specific groups of people or specific disease types. This could enhance the model's general fairness and accuracy. In summary, a large and varied dataset is essential for predicting diseases because it helps to improve accuracy, generalization, robustness, and representation.

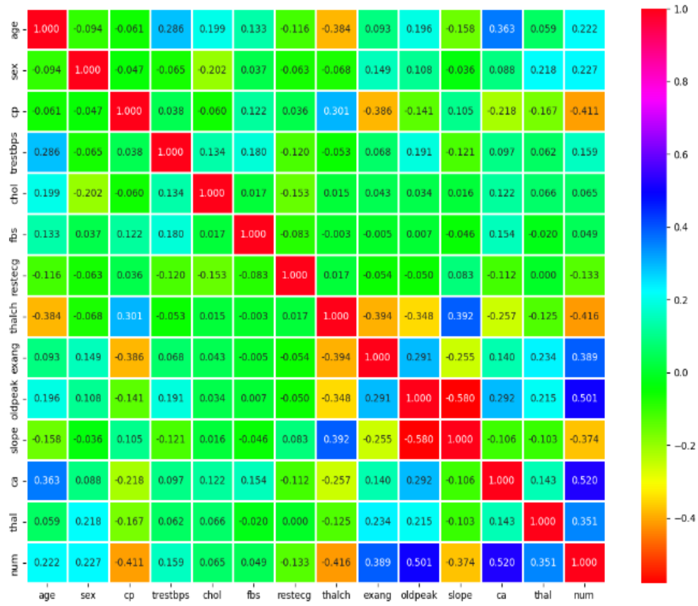
Data set Description

```
data.head()
```

	id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
0	1	63	Male	Cleveland	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	2.3	downsloping	0.0	fixed defect	0
1	2	67	Male	Cleveland	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	1.5	flat	3.0	normal	2
2	3	67	Male	Cleveland	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	2.6	flat	2.0	reversable defect	1

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 920 entries, 0 to 919
Data columns (total 16 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   id        920 non-null    int64
 1   age       920 non-null    int64
 2   sex       920 non-null    object
 3   dataset   920 non-null    object
 4   cp        920 non-null    object
 5   trestbps  861 non-null    float64
 6   chol      890 non-null    float64
 7   fbs       830 non-null    object
 8   restecg   918 non-null    object
 9   thalch    865 non-null    float64
 10  exang     865 non-null    object
 11  oldpeak   858 non-null    float64
 12  slope     611 non-null    object
 13  ca        309 non-null    float64
 14  thal      434 non-null    object
 15  num       920 non-null    int64
dtypes: float64(5), int64(3), object(8)
memory usage: 115.1+ KB
```

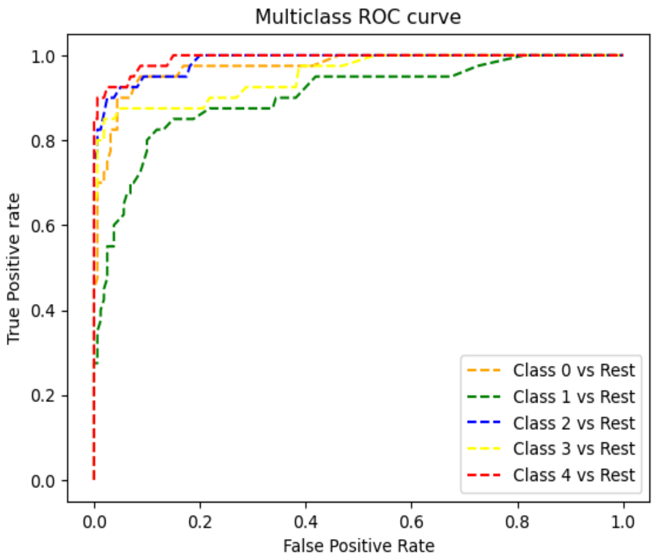
A correlation map can assist uncover significant factors and correlations between variables that may be related to the disease, which can be utilized to construct more precise predictive models.



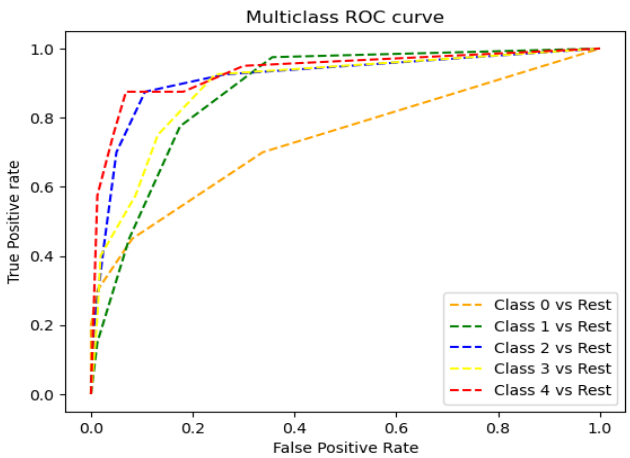
Algorithms Used:

Logistic Regression: In the context of heart disease prediction, logistic regression can be used to estimate the probability that an individual will develop heart disease based on various risk factors such as age, gender, blood pressure, cholesterol levels, and family history of heart disease. These risk factors are typically used as predictor variables in the logistic regression model. The logistic regression model estimates the probability of developing heart disease based on the values of the predictor variables. The output of the model is a probability value between 0 and 1, where 0 indicates no risk of developing heart disease and 1 indicates a high risk of developing heart disease. The model uses a threshold value (typically 0.5) to classify individuals as either having a high or low risk of developing heart disease.

Random Forest: Random Forest is a machine learning algorithm that can be used for classification tasks such as heart disease prediction. It works by constructing a multitude of decision trees and then combining their results to make a final prediction. By using multiclass ROC analysis, the performance of the random forest model can be evaluated across all classes, which can provide insight into the model's ability to correctly classify individuals with different levels of heart disease severity. This information can be used to refine the model and improve its performance for heart disease prediction.



K-NN: K-NN (K-Nearest Neighbors) is a machine learning algorithm that can also be used for classification tasks such as heart disease prediction. The K-NN algorithm works by identifying the K nearest neighbors to a new data point and using their labels to make a prediction. To evaluate the performance of a K-NN model for heart disease prediction, multiclass ROC analysis can also be used. The multiclass ROC curve is a graphical representation of the trade-off between the true positive rate (sensitivity) and false positive rate (1-specificity) for each class at different threshold values.



Linear SVM: In the context of heart disease prediction, linear SVM can be used to classify individuals into one of several classes based on their risk factors. The SVM algorithm finds the optimal hyperplane that separates the different classes in the dataset based on the support vectors. To build a linear SVM model for heart disease prediction, a labeled dataset is typically used to train the model. The model learns the relationships between the predictor variables and the outcome variable (heart disease) from the training data and then uses this knowledge to predict the likelihood of heart disease in new individuals.

Naive Bayes: Naive Bayes is a probabilistic machine learning algorithm that can be used for classification tasks, including the prediction of heart disease. In order to make predictions, the Naive Bayes algorithm requires a set of training data that includes information about individuals who have and have not been diagnosed with heart disease. Once the training data has been processed, the Naive Bayes algorithm can be used to predict the probability of an individual having heart disease based on their attributes. The algorithm calculates the probability of each attribute given the presence or absence of heart disease, and combines these probabilities using Bayes' theorem to calculate the overall probability of heart disease given the person's attributes.

Efficiency Table

Table-3

S.No	Classifier	Accuracy
1	Logistic Regression	0.4300
2	K-NN	0.8703
3	Random Forest	0.9640
4	Linear SVM	0.3650
5	Naive Bayes	0.5050

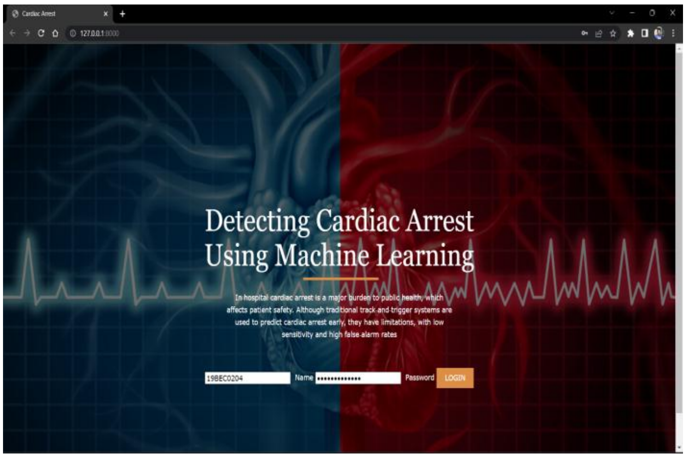
Train and Test Accuracy:

By training and testing the data models we get the train and test accuracy of the data.

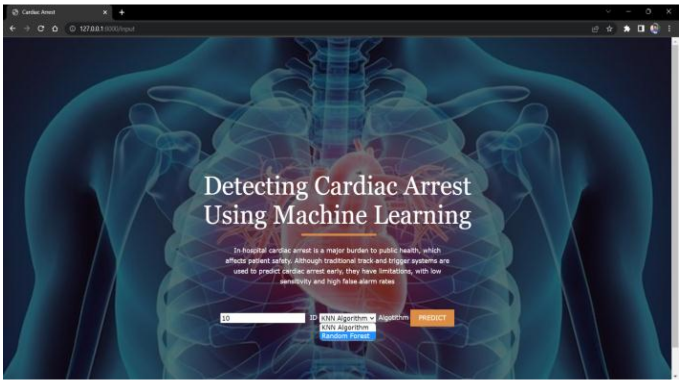
	model	Classifier	Train-Accuracy	Test-Accuracy
0	LR	Logisticregression	0.4983	0.4300
1	KNN	KNeighborsClassifier	0.9657	0.8703
2	Random Forest	RandomForestClassifier	1.0000	0.9640
3	Linear SVM	SGDClassifier	0.3830	0.3650
4	Naive Bayes	GaussianNB	0.5566	0.5050

Disease Prediction:

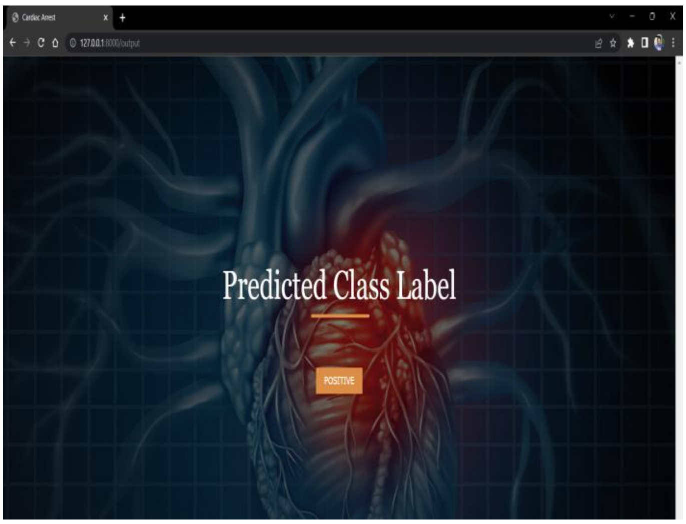
Choosing the Most Efficient algorithm:



Both K-NN and Random Forest can be effective in predicting heart disease because they can handle complex and non-linear relationships between the features and the target variable. Additionally, they can handle missing data and outliers effectively.



Predicting if the patient is suffering from disease or not



CONCLUSION

In conclusion, our heart monitoring system, comprising both hardware and software components, represents a significant advancement in predicting the risk of heart disease. By utilizing a range of sensors for data collection, followed by thorough preprocessing and training of machine learning models, we have developed a highly accurate prediction tool. Deploying this trained model on a website enables users to conveniently input their health information and obtain personalized risk assessments. The accuracy of the classifiers which we have used K-NN and Random Forest is 0.8703 and 0.9640 so by this we can conclude that the disease prediction is very accurate and there is a very little room for error and its potential to feed more data can only increase its efficiency furthermore and holds promise for improving overall cardiovascular health outcomes.

REFERENCES

[1] Riyazulla Rahman.J,Shridhar Sanshi and N. Nasurudeen Ahamed, "Health Monitoring and Predicting System using Internet of Things & Machine Learning" 7th ICACCS,2021.

[2] Dr.Yogesh Kumar Sharma ,Khatal Sunil S , "Health Care Patient Monitoring using IoT and Machine Learning", IOSR Journal of Engineering, 2019.

[3] Dr.Jayant Shekhar, Desalegn Abebaw, "Temperature and Heart Attack Detection using IOT(Arduino and ThingSpeak)", IJACST Volume 7 No.11, November 2018.

[4] Sahana S Khamitkar, "IoT based System for Heart Rate Monitoring", IJERT Vol. 9 Issue 07, July-2020.

[5] Helton Hugo de Carvalho Junior , Robson Luiz Moreno, " A heart disease recognition embedded system with fuzzy cluster algorithm", Elsevierhealth, 2013.

[6] Reetu Singh, "A Review on Heart Disease Prediction using Unsupervised and Supervised Learning", IJARESM Volume 8, Issue 7, July-2020.

[7] A. S. Thanuja Nishadi, "Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms By Python Jupyterlab", IJARP Volume 3 Issue 8, August 2019.

[8] Madhumita Pal, Smita Parija, "Prediction of Heart Diseases using Random Forest" , ICCIEA 2020.

[9] Harshit Jindal, Sarthak Agrawal, "Heart disease prediction using machine learning Algorithms", ICCRDA 2020.

[10] B. Manoj Kumar, P S.Uma Priyadarsini, "Efficient Prediction of Heart Disease using SVM Classification Algorithm and Compare its Performance with Linear

Regression in Terms of Accuracy", Journal of Pharmaceutical Negative Results ,Volume 13, Special Issue 4 2022.

[11] M. Marimuthu, M. Abinaya and M. Abinaya, " A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach", International Journal of Computer Applications Volume181, No.18, September 2018.

[12] Arun.R, N.Deepa, "Heart disease prediction system using naive bayes", International Journal of Pure and Applied Mathematics, Volume 119 No. 16, 2018,

[13] Kompella Sri Charan, Kolluru S S N S Mahendranath, "Heart Disease Prediction Using Random Forest Algorithm", IRJET, Volume: 09 Issue: 03 Mar 2022.

[14] Akkem Yaganteeswarudu, "Multi Disease Prediction Model by using Machine Learning and Flask API", ICCES 2020.

BIOGRAPHIES

Dangeti Sai Chaithrik
dsaichaithrik@gmail.com

Chiluka Harshith Reddy
harshithreddy382@gmail.com