DESIGN AND DEVELOPMENT OF IOT BASED HEALTHCARE MONITORING DEVICE FOR EARLY DIAGNOSIS OF HEART DISEASE USING AI

A thesis submitted in partial fulfilment of the requirements for the award of the degree of

B.Tech.

in

Instrumentation and Control Engineering

By

Hemangani N 110118031

Akshyah 110118006

DEPARTMENT OF

INSTRUMENTATION AND CONTROL ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

TIRUCHIRAPPALLI – 620015

MAY-2022

BONAFIDE CERTIFICATE

This is to certify that the project titled **DESIGN AND DEVELOPMENT OF IOT BASED HEALTHCARE MONITORING DEVICE FOR EARLY DIAGNOSIS OF HEART DISEASE USING AI** is a bonafide record of the work done by

Hemangani N	**110118031**
Akshyah	**110118006**

in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Instrumentation and Control Engineering** of the **NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI**, during the year 2021-2022.

Dr. C GEETHA
Guide

for **Dr. K. DHANALAKSHMI** 17/5
Head of the Department

ABSTRACT

The aim of the project is to develop an IoT healthcare monitoring & heart disease prediction device using AI. People all over the world are prone to chronic illnesses like chronic respiratory disease, heart disease, and diabetes. Among them, heart disease with disparate features or symptoms complicates diagnosis. Because of the emergence of smart wearable gadgets and the "Internet of Things" (IoT), solutions have become necessary for diagnosis. The hardware components (sensors) collect data from different patients. The heart feature extraction from signals is done to get significant features.

Furthermore, the feature extraction of other attributes is also gathered. All these features are collected and subjected to the diagnostic system using Machine Learning, Deep Learning, and artificial intelligence techniques. Here we are planning to design a hand wearable device. IoT-based healthcare systems and to provide a systematic review of its enabling technologies, services, and applications. IoT technology has helped healthcare professionals to monitor and diagnose several health issues, measure many health parameters, and provide diagnostic facilities at remote locations.

The major disease caused by human death nowadays is heart disease, due to it happening suddenly and without significant symptoms, leading patients to miss the best time for first aid. With the development of IoT technology combined with the healthcare industry. It is providing technical support for clinic staff to predict and monitor heart disease patients remotely. Furthermore, compare the different proposed method's performance and present the best framework for heart disease continuous prediction and monitoring.

ACKNOWLEDGEMENTS

We would like to thank the following people for their support and guidance, without whom the completion of this project in fruition would not be possible.

We wish to express our sincere thanks to our research guide, Dr. C. Geetha, for providing us with all the expertise and knowledge for the implementation of our project work and for having faith in our aptitude through this entire period.

Dr. K.Dhanalakshmi, the Head of the Department of Instrumentation and Control Engineering for providing us with the support to finish the project.

We would also like to extend our gratitude to all the project committee members: Dr. C.Geetha, Mr. Goldin R Bennet, and to all the faculty and staff members of the Department of Instrumentation and Control Engineering for their support with our requirements for lab time.

Our reviewers, Dr. Shiraz Sohail, and Dr. B Janet for their valuable insights and advice provided during the review sessions.

We would also like to thank our parents and friends for their constant support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Healthcare Monitoring Overview

Healthcare monitoring systems in hospitals and many other health centres have experienced significant growth, and portable healthcare monitoring systems with emerging technologies are becoming of great concern to many countries worldwide nowadays. The advent of Internet of Things (IoT) technologies facilitates the progress of healthcare from face-to-face consulting to telemedicine. We propose a smart healthcare system in an IoT environment that can monitor a patient's basic health signs as well as the room condition where the patients are now in real-time.

1.2 Motivation

Nowadays, heart disease is the leading cause of death worldwide. Predicting heart disease is a complex task since it requires experience along with advanced knowledge. Internet of Things (IoT) technology has lately been adopted in healthcare systems to collect sensor values for heart disease diagnosis and prediction. These hybrid technologies work smartly to improve the decision-making process. ML empowers the IoT to demystify hidden patterns in bulk data for optimal prediction.

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5 CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 6 features that can be used to predict possible heart disease. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia, or already established disease) need early detection and management wherein a machine learning model can be of great help.

1.3 Objectives

This project aims at designing and developing an IOT based healthcare monitoring and Heart Disease Prediction device using AI and the sub objectives are:

1. To design and develop a wearable IOT healthcare monitoring system.
2. Successful Data acquisition and transfer from the sensors.
3. To process the data acquired and develop a machine learning model for diagnosing Heart Disease and other ailments using AI.
4. To integrate the hardware with software and analyse and validate the system.
5. The further improvements in the project include making Web/app user interface design and implementation for better understanding and visualisation purposes.

1.4 Significant contribution

This project directly contributes to improving healthcare monitoring and early diagnosing by performing data analysis on the collected health records, extracting various features from the collected data, and using those features to train the ML models and use it for prediction and live monitoring.

1.5 Outline of the thesis

1. In chapter 2, the methodology of how the project works will be discussed
2. In chapter 3, the proposed and implemented system will be discussed
3. In chapter 4, the results of the research will be discussed
4. In chapter 5, the summary, conclusion and future scope of the research will be discussed

CHAPTER 2

LITERATURE SURVEY

Bo Jin, Chao Che et al. (2018) proposed a “Predicting the Risk of Heart Failure With EHR Sequential Data Modeling” model designed by applying neural networks. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to use one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimising the essential principles of an extended memory network model. By analysing the results, we tend to reveal the importance of respecting the sequential nature of clinical records [1].

Aakash Chauhan et al. (2018) presented “Heart Disease Prediction using Evolutionary Rule Learning”. This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on the patient's dataset. This will facilitate (help) in decreasing the amount of services and show that the overwhelming majority of the rules helps within the best prediction of coronary sickness [2].

Ashir Javeed, Shijie Zhou et al. (2017) designed “An Intelligent Learning System based on Random Search Algorithm and Optimised Random Forest Model for Improved Heart Disease Detection”. This paper uses a random search algorithm (RSA) for factor selection and random forest model for diagnosing cardiovascular disease. This model is principally optimised for using grid search algorithmic programs.

Two forms of experiments are used for cardiovascular disease prediction. In the first form, only a random forest model is developed and within the second experiment the proposed Random Search Algorithm based random forest model is developed. This methodology is efficient and less complex than conventional random forest models. Compared to conventional random forest it produces 3.3% higher accuracy. The proposed learning system can help the physicians to improve the quality of heart failure detection [3].

"Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" proposed by Senthil Kumar Mohan, Chandrasegar Thirumalai et al. (2019) was an efficient technique using hybrid machine learning methodology. The hybrid approach is combination of random forest and linear method. The dataset and subsets of attributes were collected for prediction. The subset of some attributes were chosen from the pre-processed knowledge(data) set of cardiovascular disease .After prep-processing , the hybrid techniques were applied and diagnosed cardiovascular disease [4].

K.Prasanna Lakshmi, Dr. C.R.K.Reddy (2015) designed "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining". In the proposed Stream Associative Classification Heart Disease Prediction (SACHDP), we used associative classification mining over landmark window of data streams. This paper contains two phases: one is generating rules from associative classification mining and next one is pruning the rules using chi-square testing and arranging the rules in an order to form a classifier. Using these phases to predict heart disease easily [5].

M.Satish, et al. (2015) used different Data Mining techniques like Rule based, Decision Tree, Naive Bayes, and Artificial Neural Network. An efficient approach called pruning classification association rule (PCAR) was used to generate association rules from cardiovascular disease warehouses for prediction of Heart Disease. Heart attack data warehouse was used for pre-processing for mining. All the above discussed data mining techniques were described [6].

Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik (2015) “An Intelligent Decision Support System for Cardiac Disease Detection”, designed a cost efficient model by using genetic algorithm optimizer technique. The weights were optimised and fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of GA and neural networks [7].

“Prediction and Diagnosis of Heart Disease by Data Mining Techniques” designed by Boshra Bahrami, Mirsaeid Hosseini Shirvani. This paper uses various classification methods for diagnosing cardiovascular disease. Classifiers like KNN, SVO classifier and Decision Tree are used to divide the datasets. Once the classification and performance evaluation the Decision tree is examined as the best one for cardiovascular disease prediction from the dataset[8].

Mamatha Alex P and Shaicy P Shaji (2019) designed “Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique”. This paper uses techniques of Artificial Neural Network, KNN, Random Forest and Support Vector Machine. Comparing the above mentioned classification techniques in data mining to predict the higher accuracy for diagnosing heart disease is Artificial Neural Network[9].

CHAPTER 3

METHODOLOGY

3.1 Approach for the proposed project

The approach followed for the proposed project is as follows:

- Literature survey:
    - To learn and understand the previous work done in the related area and the feasibility of our idea and implementation.
    - To understand the concept of heart diseases and the vital signs that affect them, for a better selection of features for the AI model.
- Building the wearable sensors:
    - Sensors to be used - body movement sensor (mems ADXL), pressure sensor (MPX10DP), temperature (LM35), heartbeat (SPO2), GPS (VK-16E), GSM/GPRS, humidity sensor, wi-fi (esp8266).
    - Selection of components and design specifications.
    - Calibration and testing of individual sensors.
- Collect Data:
    - Data acquisition from sensors.
    - Collection of datasets from online sources (UCI machine learning repository, Hungarian heart disease dataset, Framingham, and Public Health).

- Build a AI model:
    - Preprocessing of the data (data cleaning, feature selection, and extraction).
    - Design of suitable neural network architecture.
    - Proper parameters selection and optimization were employed.
    - Training, testing, and debugging of the model.

3.2 Flowcharts of the system

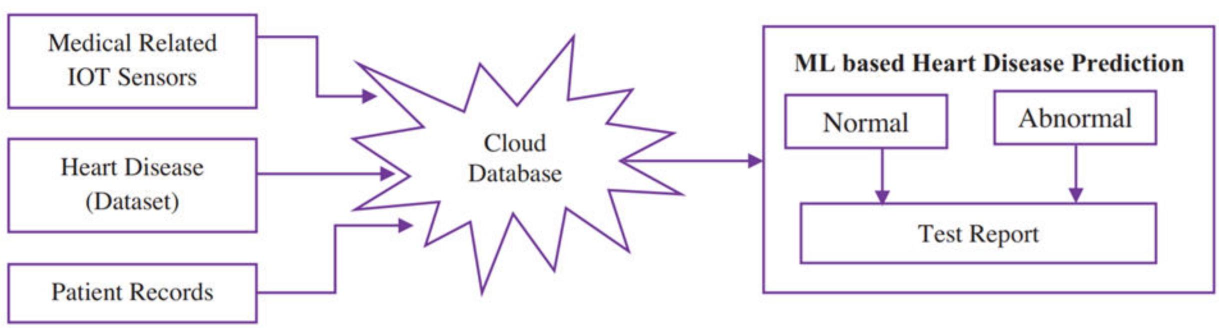

Figure 3.1 : Overall flowchart

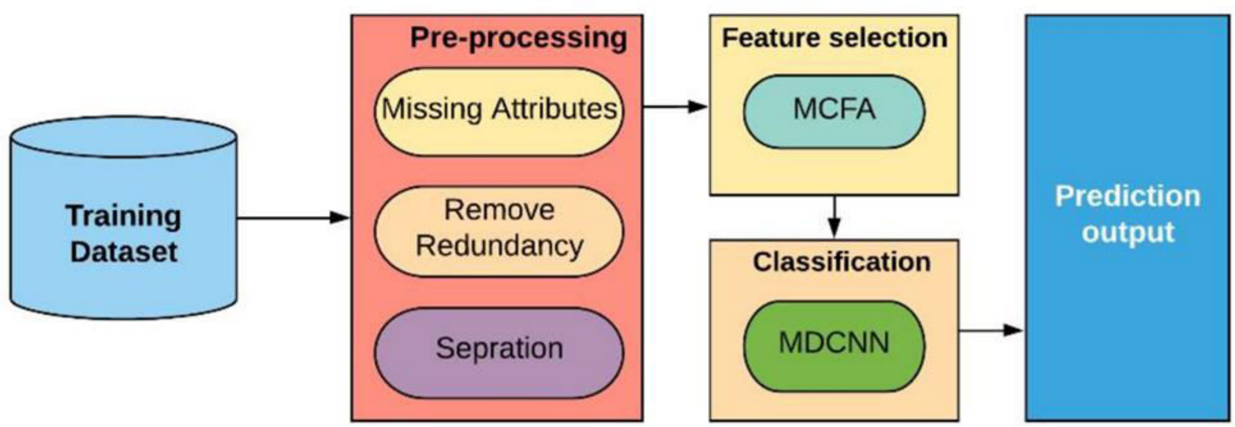

Figure 3.2 : ML flowchart

3.3 Data Acquisition and Data Analysis

3.3.1 Attribute Information

1. Age: age of the patient [years]
2. Sex: sex of the patient [M: Male, F: Female]
3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
4. RestingBP: resting blood pressure [mm Hg]
5. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
6. ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
7. HeartDisease: output class [1: heart disease, 0: Normal]

This dataset was created by combining different datasets already available independently but not combined before. In this dataset, 5 heart datasets are combined over 6 common features which makes it the largest heart disease dataset available so far for research purposes.

The five datasets used for its curation are:

1. Cleveland: 303 observations
2. Hungarian: 294 observations
3. Switzerland: 123 observations
4. Long Beach VA: 200 observations
5. Stalog (Heart) Data Set: 270 observations

Total: 1190 observations

Duplicated: 272 observations

Final dataset: 918 observations

Every dataset used can be found under the Index of heart disease datasets from UCI Machine Learning Repository on the following link:

https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/

Acknowledgments:

Creators:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Centre, Long Beach, and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Donor:

David W. Aha (aha '@' ics.uci.edu) (714) 856-8779

This dataset is to be used for building a predictive machine learning model for early-stage heart disease detection.

3.3.2 Handling Categorical Variables

Categorical variables/features are any feature type can be classified into two major types:

- Nominal
- Ordinal

Nominal variables are variables that have two or more categories which do not have any kind of order associated with them. For example, if gender is classified into two groups, i.e. male and female, it can be considered as a nominal variable.Ordinal variables, on the other hand, have “levels” or categories with a particular order associated with them. For example, an ordinal categorical variable can be a feature with three different levels: low, medium and high. Order is important.

It is a binary classification problem: the target here is not skewed but we use the best metric for this binary classification problem which would be Area Under the ROC Curve (AUC). We can use precision and recall too, but AUC combines these two metrics. Thus, we will be using AUC to evaluate the model that we build on this dataset.

We have to know that computers do not understand text data and thus, we need to convert these categories to numbers.

For not Tree based Machine Learning Algorithms the best way to go will be to use One-Hot Encoding. One-Hot-Encoding has the advantage that the result is binary rather than ordinal and that everything sits in an orthogonal vector space. The disadvantage is that for high cardinality, the feature space can really blow up quickly and you start fighting with the curse of dimensionality. In these cases, I typically employ one-hot-encoding followed by PCA for dimensionality reduction. I find that the judicious combination of one-hot plus PCA can seldom be beaten by other encoding schemes. PCA finds the linear overlap, so will naturally tend to group similar features into the same feature.

For Tree based Machine Learning Algorithms the best way to go is with Label Encoding. LabelEncoder can turn [dog,cat,dog,mouse,cat] into [1,2,1,3,2], but then the imposed ordinality means that the average of dog and mouse is cat. Still there are algorithms like decision trees and random forests that can work with categorical variables just fine and LabelEncoder can be used to store values using less disk space.

	A	B	C	D	E	F	G	H	I	J
1	age	anaemia	high_blood_pres	sex	smoking	height	weight	Alcohol	active	diagnosis of heart disease (Target Variable)
2	75	0	1	1	0	110	80	0	1	1
3	55	0	0	1	0	140	90	0	1	1
4	65	0	0	1	1	130	70	0	0	1
5	50	1	0	1	0	50;	00	0	1	1
6	65	1	0	0	0	100	60	1	0	1
7	90	1	1	1	1	120	80	0	1	1
8	75	1	0	1	0	130	80	0	1	1
9	60	1	0	1	1	130	90	1	1	1
10	65	0	0	0	0	110	70	0	0	1
11	80	1	1	1	1	110	60	0	1	1
12	75	1	1	1	1	120	80	0	1	1
13	62	0	1	1	1	120	80	0	0	1
14	45	1	0	1	0	120	80	0	1	1
15	50	1	1	1	0	110	70	0	1	1
16	49	1	1	0	0	130	90	0	0	0
17	82	1	0	1	0	120	80	0	1	1
18	87	1	0	1	0	130	70	0	1	1
19	45	0	0	1	0	110	70	0	1	1
20	70	1	1	0	0	100	70	0	0	1
21	48	1	0	0	0	120	70	1	0	1
22	65	1	1	0	0	120	80	0	0	0
23	65	1	1	0	0	130	80	1	0	1
24	68	1	1	1	1	145	85	0	1	1
25	53	0	0	1	0	110	60	1	1	0
26	75	0	1	0	0	150	90	1	0	1
27	80	0	0	1	1	30;	00	1	1	1
28	95	1	1	0	0	130	90	0	0	1

Figure 3.3 : Processed Dataset excel snapshot

```
Rows     =  917
Columns  =  7
```

	Age	Sex	CP	BP	HR	EA	HeartDisease
0	40	M	ATA	140	172	N	0
1	49	F	NAP	160	156	N	1
2	37	M	ATA	130	98	N	0
3	48	F	ASY	138	108	Y	1
4	54	M	NAP	150	122	N	0

	count	mean	std	min	25%	50%	75%	max
Age	917.0	53.509269	9.437636	28.0	47.0	54.0	60.0	77.0
BP	917.0	132.540894	17.999749	80.0	120.0	130.0	140.0	200.0
HR	917.0	136.789531	25.467129	60.0	120.0	138.0	156.0	202.0
HeartDisease	917.0	0.552890	0.497466	0.0	0.0	1.0	1.0	1.0

Figure 3.4 : Dataset standard moments

3.4 Feature Extraction

Attributes: No zero variance and no extremely high variance.

Target Variable:

- Almost 55% of the patients had a heart disease.
- 508 patients had heart disease.
- Almost 45% of patients didn't have a heart disease.
- 410 patients didn't have a heart disease.
- Nothing much for the skewness. Quite a normal-like distribution for the numerical features.

Exploratory Data Analysis and Feature Engineering:

Figure 3.5 : Dataset distribution

Based on the gender; Men are almost 2.44 times more likely to have a heart disease than women. We can observe clear differences among the chest pain types.

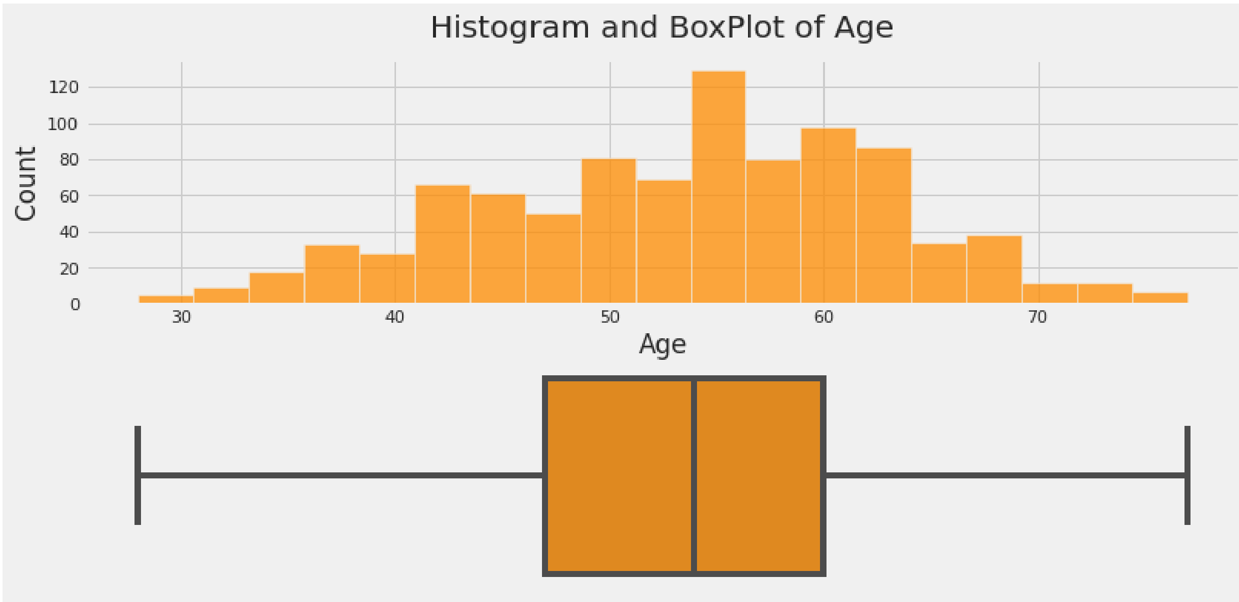

Figure 3.6 : Histogram and BoxPlot of age

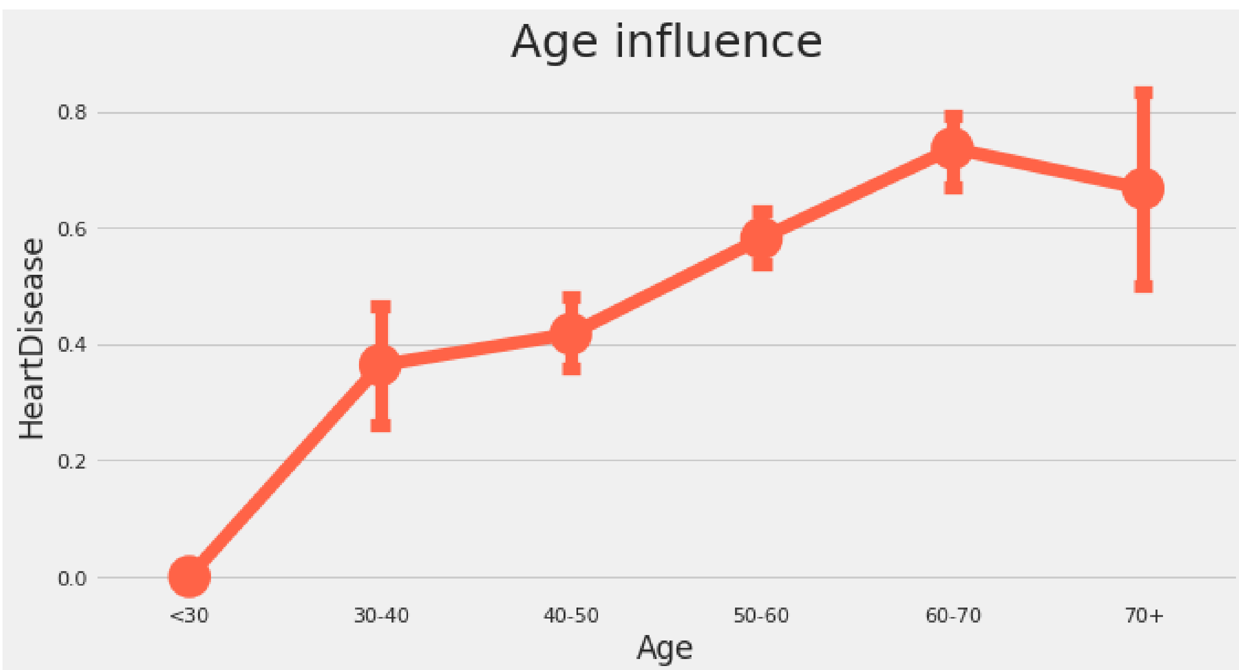

Figure 3.7 : Age influence

Conclusion (linear relationship) : as the age of patients increases, the risk of heart disease increases (This disease affects mainly the elderly).

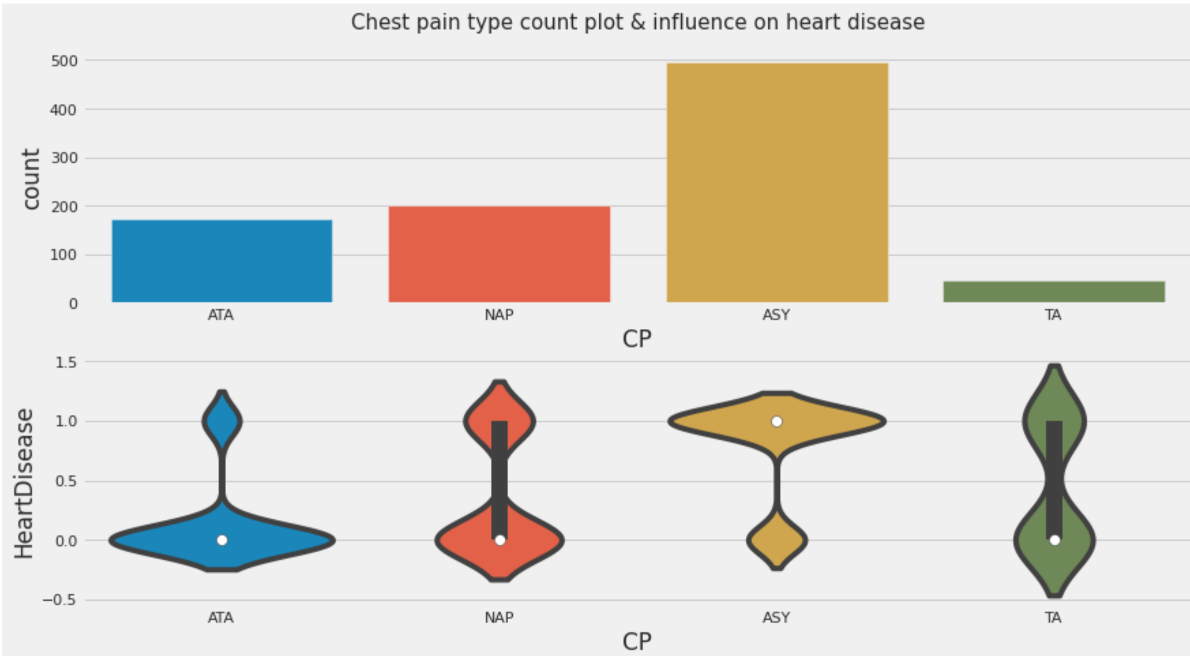

Figure 3.8 : Chest pain type and influence on heart disease

Conclusion : Asymptomacy in most cases has a stronger influence on the disease's progression, while patients with Atypical Angina have fewer (average) cases.

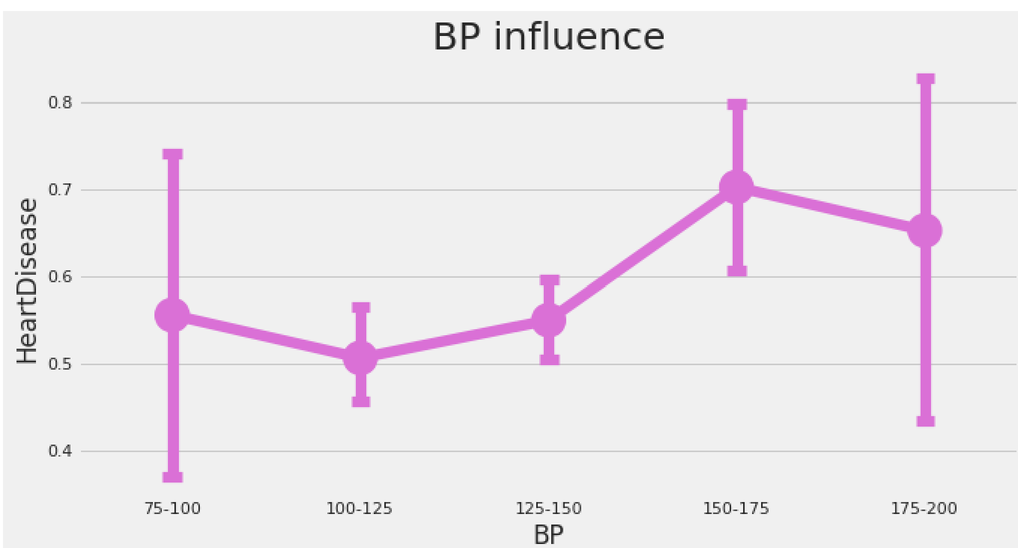

Figure 3.9 : BP influence

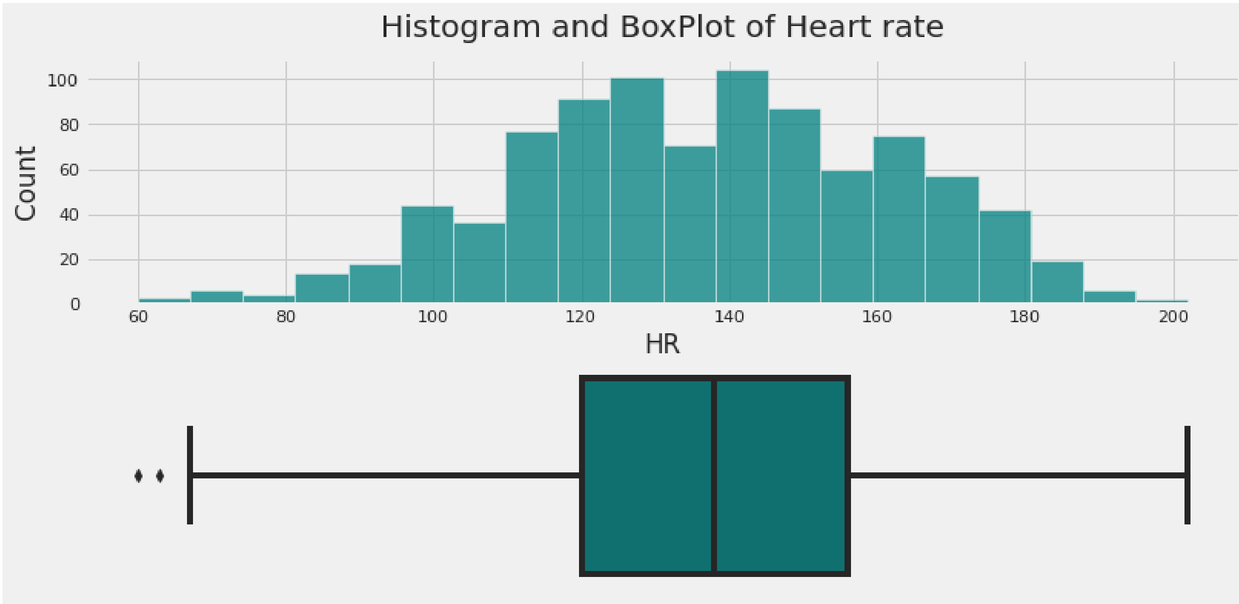

Figure 3.10 : Histogram and Boxplot of heart rate

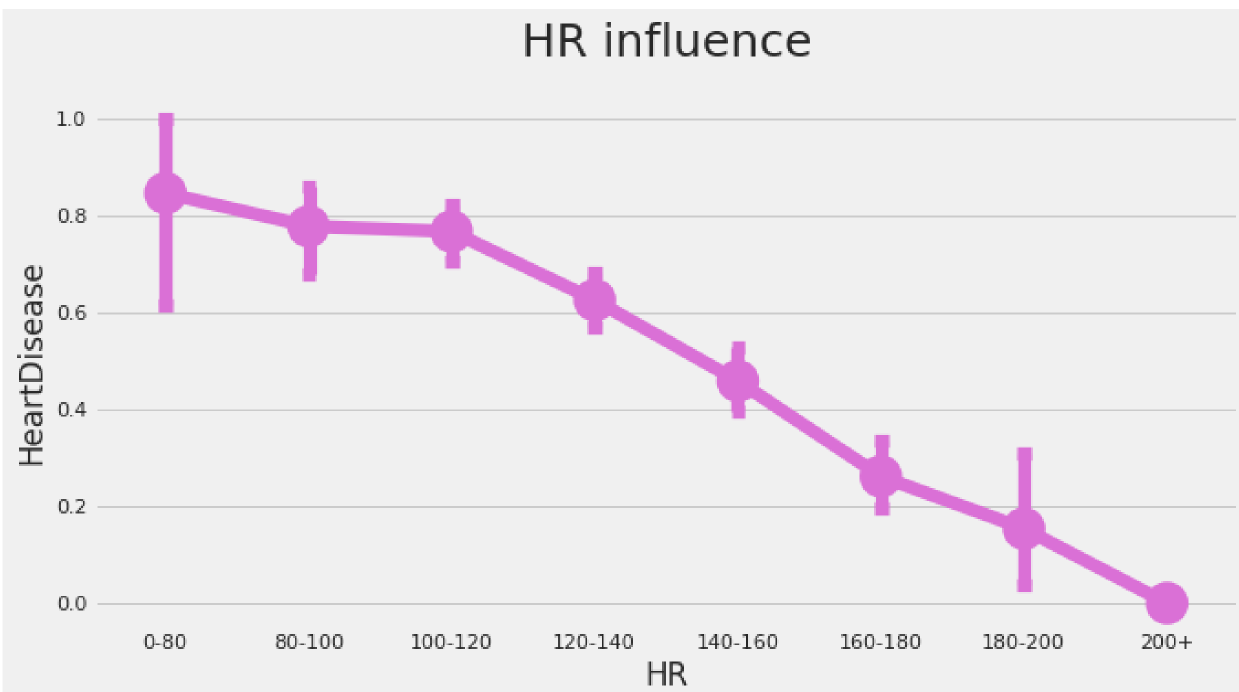

Figure 3.11 : HR influence

Explanation : It can therefore be concluded that the relationship is inverse in relation to the age of the patient To estimate your maximum age-related heart rate, subtract your age from 220.

ExerciseAngina: exercise-induced angina with 'Yes' almost 2.4 times more likely to have a heart disease than exercise-induced angina with 'No'.

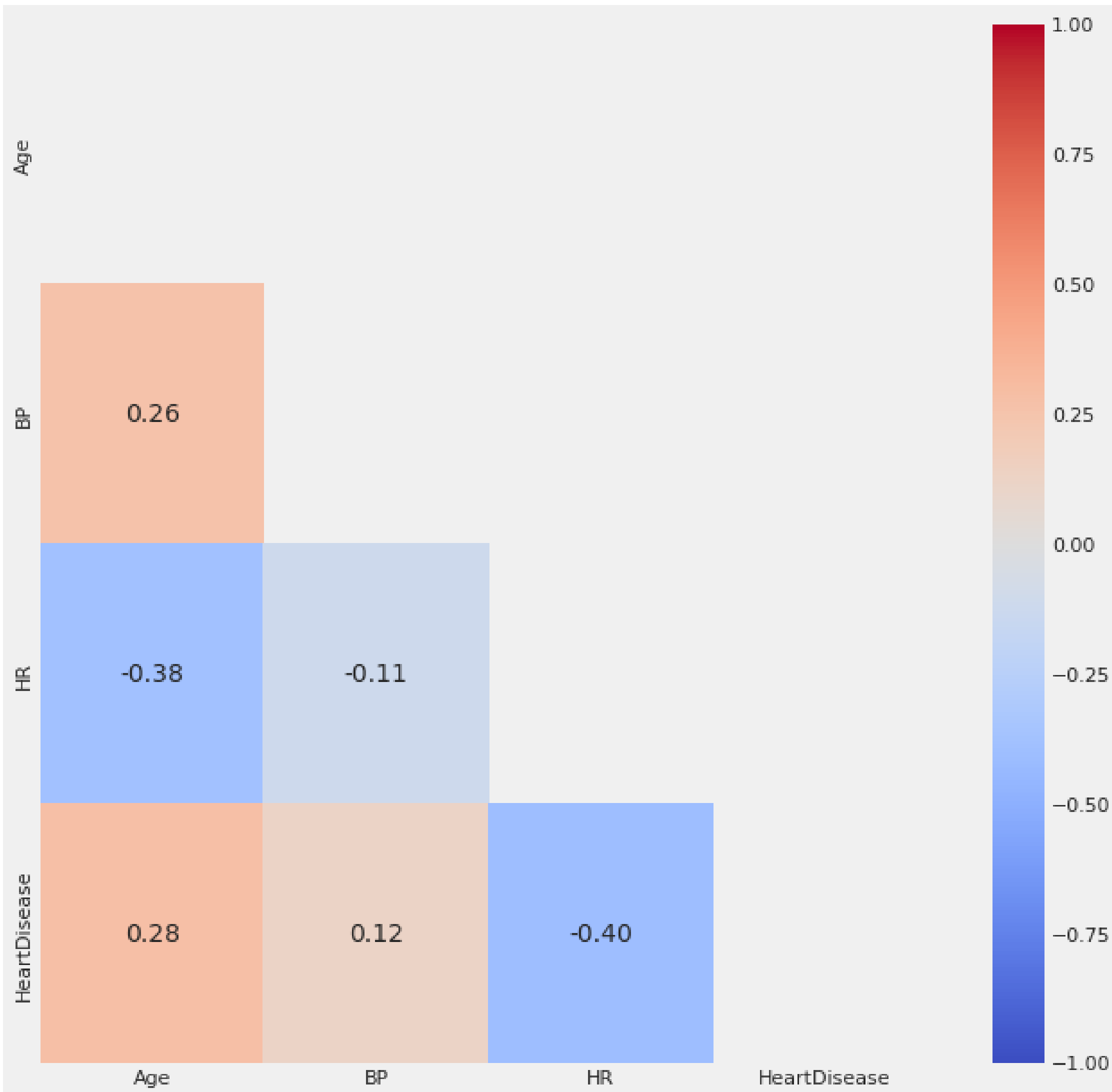

Figure 3.12 : Heatmap

Based on the matrix, we can observe weak level correlation between the numerical features and the target variable. Maximum heart rate has negative correlation with heart disease.

3.5 Machine Learning Model Construction

Machine learning is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence.

3.5.1 Problem and the metric

Based on the data and data dictionary, We have a classification problem. We will make classification on the target variable Heart Disease And we will build a model to get the best classification possible on the target variable. For that we will look at the balance of the target variable. Target variable has balanced or balanced like data For that reason we will use Accuracy score.

3.5.2 Insights from the Exploratory Data Analysis

- Target variable has close to balanced data.
- Numerical features have weak correlation with the target variable.
- Maximum heart rate has a negative correlation with heart disease.
- Based on the gender; Men are almost 2.44 times more likely to have a heart disease than women.
- We can observe clear differences among the chest pain types. Person with ASY: Asymptomatic chest pain is almost 6 times more likely to have a heart disease than person with ATA Atypical Angina chest pain.
- ExerciseAngina: exercise-induced angina with 'Yes' almost 2.4 times more likely to have a heart disease than exercise-induced angina with 'No'.

3.5.3 Model Selection

We'll use a dummy classifier model as a base model And then we will use Logistic & Linear Discriminant & KNeighbors and Support Vector Machine models with and without scaler. And then we will use ensemble models, Adaboost, Randomforest, Gradient Boosting and Extra Trees. We will use XGBoost,LightGBM & Catboost. Finally we will look in detail at hyperparameter tuning for Catboost.

Machine learning models used are as follows:

- model : DummyClassifier(constant=1, strategy='constant') and accuracy score is : 0.5942

Logistic & Linear Discriminant & SVC & KNN

- model : LogisticRegression(solver='liblinear') and accuracy score is : 0.8841
- model : LinearDiscriminantAnalysis() and accuracy score is : 0.8696
- model : SVC() and accuracy score is : 0.7246
- model : KNeighborsClassifier() and accuracy score is : 0.7174

Logistic & Linear Discriminant & SVC & KNN with Scaler

- model : LogisticRegression(solver='liblinear') and accuracy score is : 0.8804
- model : LinearDiscriminantAnalysis() and accuracy score is : 0.8696
- model : SVC() and accuracy score is : 0.8841
- model : KNeighborsClassifier() and accuracy score is : 0.8841

Ensemble Models (AdaBoost & Gradient Boosting & Random Forest & Extra Trees)

- model : AdaBoostClassifier(random_state=0) and accuracy score is : 0.8659
- model : GradientBoostingClassifier(random_state=0) and accuracy score is : 0.8768
- model : RandomForestClassifier(random_state=0) and accuracy score is : 0.8877
- model : ExtraTreesClassifier(random_state=0) and accuracy score is : 0.8804

XGBoost & LightGBM & Catboost

- XGBoost Accuracy :0.8297
- LightGBM Accuracy :0.8732
- CATBOOST Accuracy :0.8804
    - Training and applying models for the classification problems. Provides compatibility with the scikit-learn tools. The default optimised objective depends on various conditions: Logloss — The target has only two different values or the target_border parameter is not None. MultiClass — The target has more than two different values and the border_count parameter is None.

ANN (Artificial Neural Network): Neural Network score = 0.809

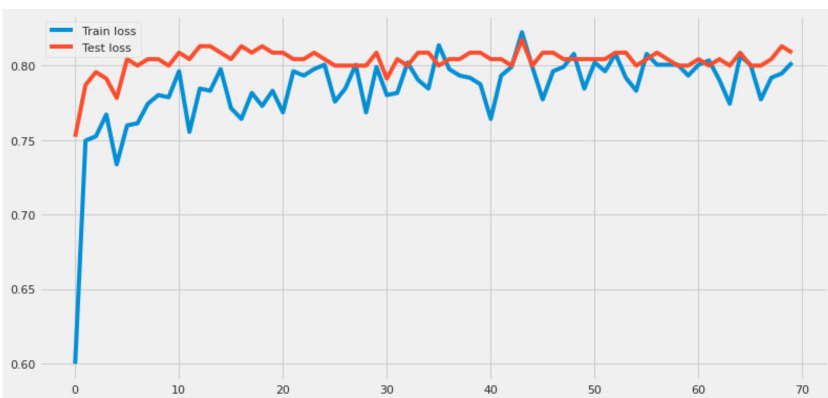

Figure 3.13 : ANN train and test loss graph

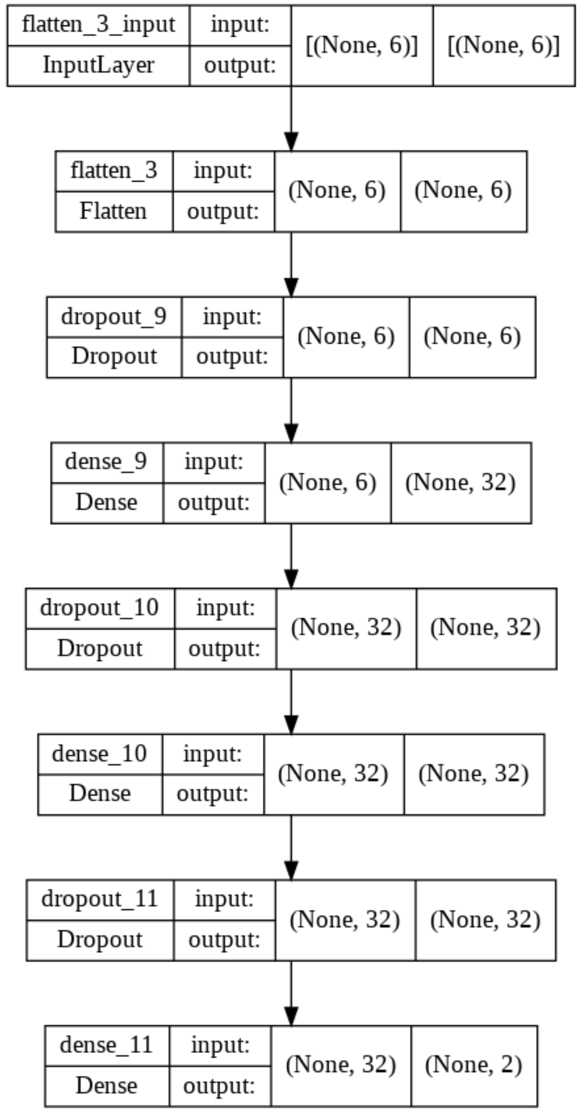

Figure 3.14 : ANN structure

Catboost HyperParameter Tuning with Optuna

Parameters:

- Objective: Supported metrics for overfitting detection and best model selection
- colsample_bylevel: this parameter speeds up the training and usually does not affect the quality.

- depth : Depth of the tree.
- boosting_type : By default, the boosting type is set to for small datasets. This prevents overfitting but it is expensive in terms of computation. Try to set the value of this parameter to to speed up the training.
- bootstrap_type : By default, the method for sampling the weights of objects is set to . The training is performed faster if the method is set and the value for the sample rate for bagging is smaller than 1.

Catboost_tuned Accuracy :0.9094. We have lifted from 0.8804 to .9094 using optimization techniques.

Summary:

- Developed model to classify heart disease cases
- Made the detailed exploratory analysis
- Decided which metric to use
- Analysed both target and features in detail
- Looked at the feature importance
- Selected the best one for the problem on hand
- Made hyperparameter tuning of the Catboost with Optuna to see the improvement

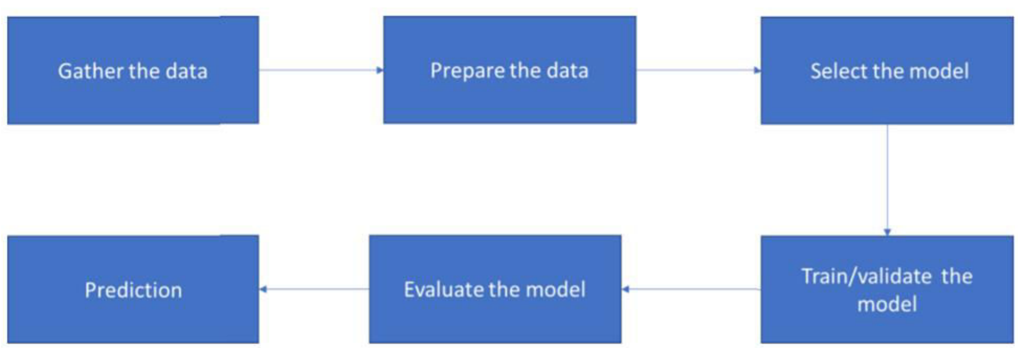

Figure 3.15 : Flow of data

CHAPTER 4

IMPLEMENTATION

4.1 Sensors

4.1.1 Heart Beat Pulse sensor module

Figure 4.1 : Heartbeat sensor module

Pulse sensor is used to test the heart rate. Sensors can be worn on the finger or earlobe and can be connected with an Arduino line via the internet. It also has an open-source app program, which can put the heart rate in real time displayed by the graph. Essence is an integrated optical heart rate sensor amplifier and noise elimination circuit. The Pulse Sensor is a well-designed low-power plug-and-play heart-rate sensor for the Arduino. It can be used to incorporate live heart-rate data into the project. And the best part is that this sensor plugs right into Arduino and easily clips onto a fingertip or earlobe. It is also super small (button-shaped) with holes, so it can be sewn into fabric.

4.1.1.1 Hardware Overview

The front of the sensor is the side with the heart logo. This is where we place our fingers. On the front side there is a small round hole, from where the King Bright's reverse mounted green LED shines.

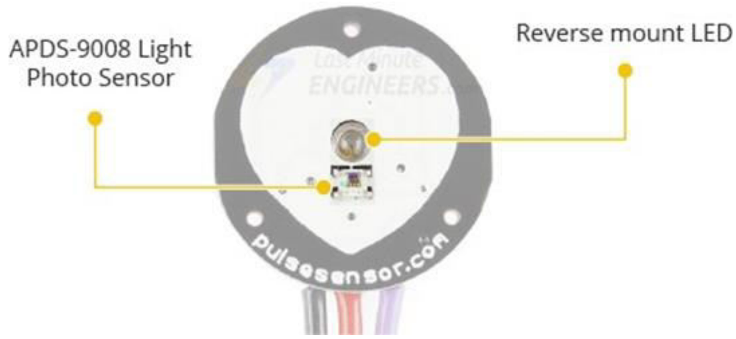

Figure 4.2 : Pulse sensor front view

Just below the LED is a small ambient light photo sensor – APDS-9008 from Avago, similar to that used in cell phones, tablets and laptops, to adjust the screen brightness in different light conditions.

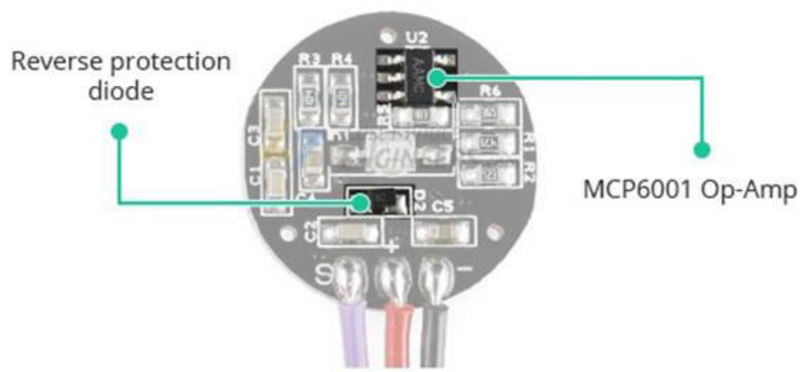

Figure 4.3 : Pulse sensor back view

On the back of the module is the rest of the components including a microchip's MCP6001 Op-Amp and a bunch of resistors and capacitors that make up the R/C filter network. There is also a reverse protection diode to prevent damage if the power leads are accidentally reversed. The module operates from a 3.3 to 5V DC Voltage supply with an operating current of < 4mA.

4.1.1.2 Technical specifications

Maximum Ratings	VCC	3.0 – 5.5V
	IMax (Maximum Current Draw)	< 4mA
	VOut (Output Voltage Range)	0.3V to Vcc
Wavelength	LED Output	565 nm
	Sensor Input	525 nm
Dimensions	L x W (PCB)	15.8mm (0.625″)
	Lead Length	20cm (7.8″)

4.1.1.3 Pulse Sensor Working

A pulse sensor or any optical heart-rate sensor, for that matter, works by shining a green light (~ 550 nm) on the finger and measuring the amount of reflected light using a photosensor. This method of pulse detection through light is called Photoplethysmogram.

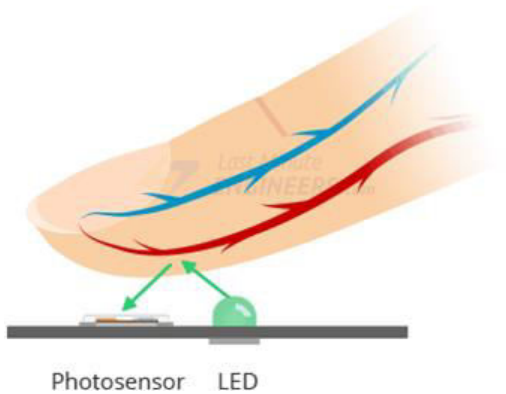

Figure 4.4 : Working diagram

The oxygenated haemoglobin in the arterial blood has the characteristic of absorbing green light. The redder the blood (the higher the haemoglobin), the more green light is absorbed. As the blood is pumped through the finger with each heartbeat, the amount of reflected light changes, creating a changing waveform at the output of the photosensor.

As we continue to shine light and take photo sensor readings, we quickly start to get a heart-beat pulse reading. This signal from the photosensor is generally small and noisy, therefore the signal is passed through an R/C filter network and then amplified using an Op Amp to create a signal that is much larger, cleaner and easier to detect.

4.1.1.4 Pulse Sensor Pinout

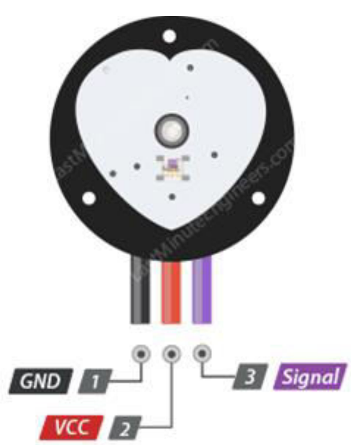

Figure 4.5 : Pulse sensor pinout

The sensor comes with a 24” flat ribbon cable with 3 male header connectors. The following diagram shows the pinout.

S (Signal) is the signal output. Connects to analog input of an Arduino.

+ (VCC) is the VCC pin. Connects to 3.3 or 5V.

– (GND) is the Ground pin.

Wiring Pulse Sensor with Arduino:

The module can be powered from 3.3 or 5V. The positive voltage connects to ‘+’ and ground connects to ‘-‘. The 3rd ‘S’ wire is the analog signal output from the sensor and this will connect to the A0 analog input of an Arduino.

4.1.2 ESP2866

The ESP8266 is a low-cost Wi-Fi microchip, with a full TCP/IP stack and microcontroller capability. We can use ESP8266 in the projects for communication. It has 3 different modes. Now we will connect ESP8266 to Arduino Uno R3 and then we will set up our microchip with AT commands. ESP8266 can be used in 2 different styles. In the first option we can use it as a receiver or transmitter. Second option is using ESP8266 as a microcontroller.

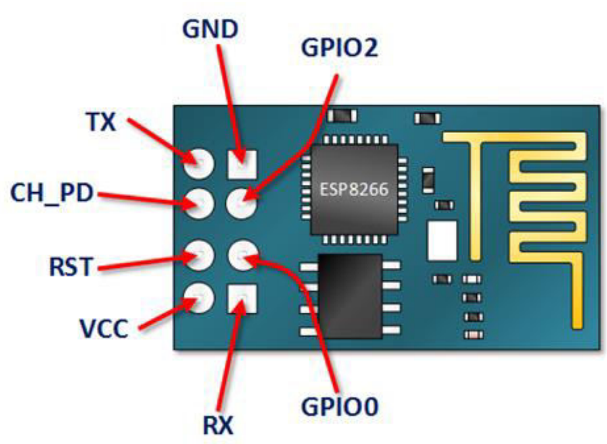

Figure 4.6 : ESP8266

We will use just 5 of these 8 pins. RX and TX are our transmitter and receiver pins. We'll use them for serial communication. Last 3 pins are about the power system of ESP8266 which are GND, VCC, and CH_PD (CH_EN). As we know that GND and VCC pins are ground and voltage pins. We will use CH_PD (CH_EN) pins for enabling our microchip. Also, we can understand the right direction of the microchip by using the golden line at the right side of it.

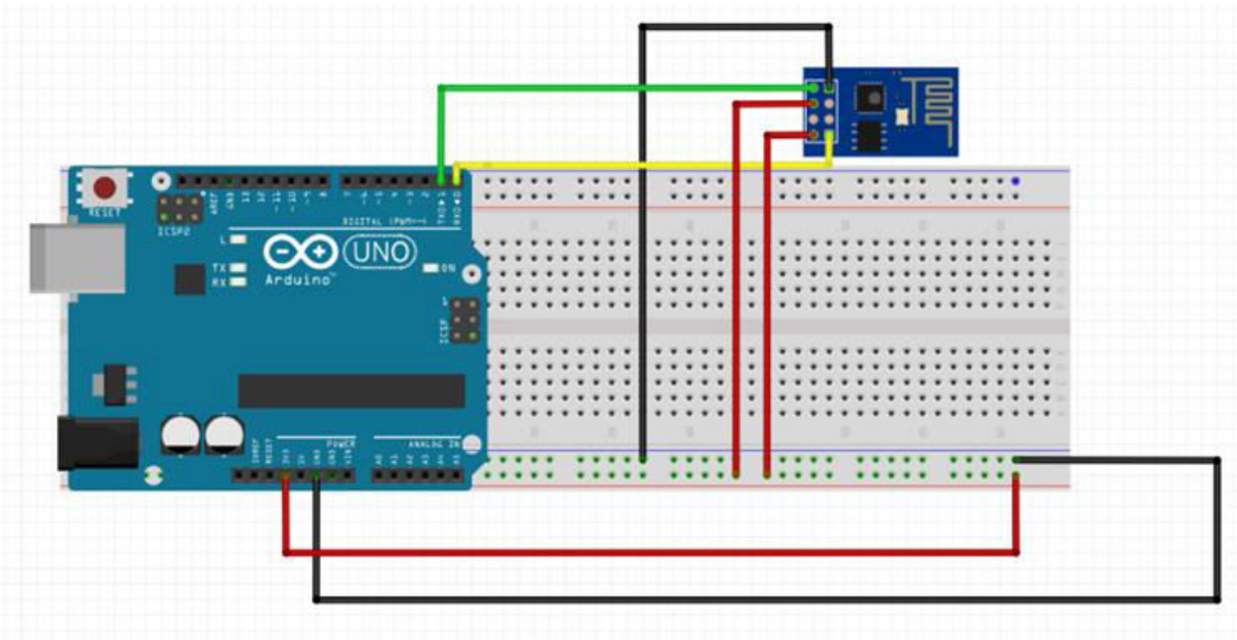

Figure 4.7 : ESP8266 wiring

We have to use a serial monitor at 115200 baud rate because this rate is the default option for ESP8266. Secondly, in the line options we should use Both NL & CR. If we do these steps correctly, we are ready to start to set up.

As the first step of setup, we should type "AT" in the serial monitor screen. Then we should get an "OK" response from this screen. If we get this, we will have an available ESP8266. Then we will type "AT+GMR" command to reach our version information.

Basic	
Command	**Description**
AT	Test AT startup
AT+RST	Restart module
AT+GMR	View version info
AT+GSLP	Enter deep-sleep mode
ATE	AT commands echo or not
AT+RESTORE	Factory Reset
AT+UART	UART configuration, [@deprecated]
AT+UART_CUR	UART current configuration
AT+UART_DEF	UART default configuration, save to flash
AT+SLEEP	Sleep mode
AT+RFPOWER	Set maximum value of RF TX Power
AT+RFVDD	Set RF TX Power according to VDD33

Figure 4.8 : AT commands table for ESP8266

We can see that we all work at 115200 baud rate. This baud rate is too fast for us and we might want to change it to 9600 which is the optimum baud rate that is used by other components. For this change we will use the "AT+UART_DEF" command which can be seen on the table. But we will have some changes on this command. This command changes default UART configurations of our device. So, we should give some configuration parameters.

"AT+UART_DEF=9600,8,1,0,0" we will use this command to change baud rate of ESP8266.

Commands	Description	Type
AT+RST	restart module	basic
AT+CWMODE	wifi mode	wifi
AT+CWJAP	join AP	wifi
AT+CWLAP	list AP	wif
AT+CWQAP	quit AP	wifi
AT+CIPSTATUS	get status	TCP/IP
AT+CIPSTART	set up TCP or UDP	TCP/IP
AT+CIPSEND	send data	TCP/IP
AT+CIPCLOSE	close TCP or UDP	TCP/IP
AT+CIFSR	get IP	TCP/IP
AT+CIPMUX	set multiple connections	TCP/IP
AT+CIPSERVER	set as server	TCP/IP

Figure 4.9 : AT commands for wifi configuration

The last step of the ESP8266 setup is about WiFi connection. In the table we can see AT commands for WiFi configuration. Usage of ESP8266 we need to know these commands for WiFi communication. The first command "AT+RST" is resetting our device. It is not changing settings, just reset the device.

"AT+CWMODE" this command is allowing us to change our device's usage mode. This command takes 1 parameter which should be integer.

"AT+CWMODE=mode" mode is representing the parameters and it looks like:

Modes:

1- Station mode (client)

2- AP mode (host)

3- AP and Station mode

That means if we want to use ESP8266 as a client of our system, we should use the "AT+CWMODE=1" command for it.

Third command is about WiFi spots. With this command we can scan WiFi spots which are located nearby. "AT+CWLAP" can list all available WiFi spots for us.

The next step of setup is joining a WiFi. I did set my device's mode as station mode then I listed all available WiFi spots. Now we will join WiFi without password. For this step we will use the "AT+CWJAP" command. But we know that for joining a WiFi we need SSID and Password. We can see all available SSIDs when we scan WiFi. Now we should change this command to "AT+CWJAP="YOUR_SSID_NAME","YOUR_PASSWORD".

We will use the "AT+CIFSR" command to find our IP. We can use this IP to communicate to our device from another system. Also, this command returns us the MAC Address of ESP8266.

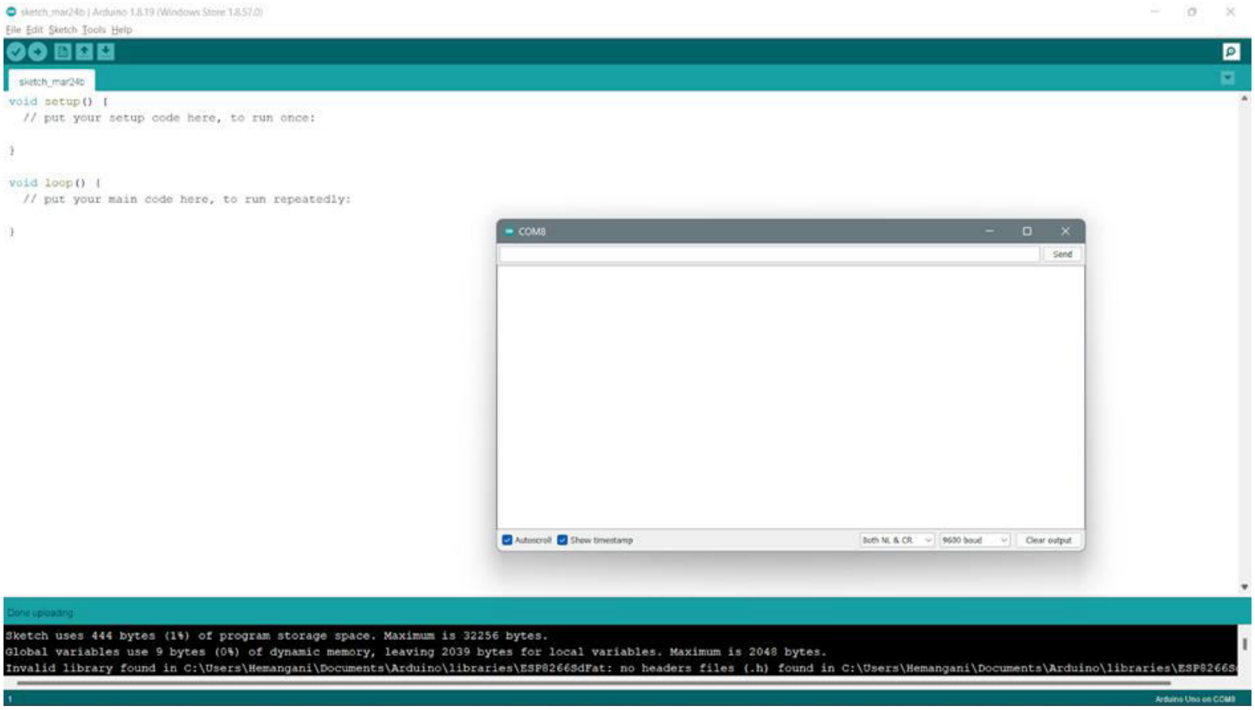

Figure 4.10 : IDE snapshot

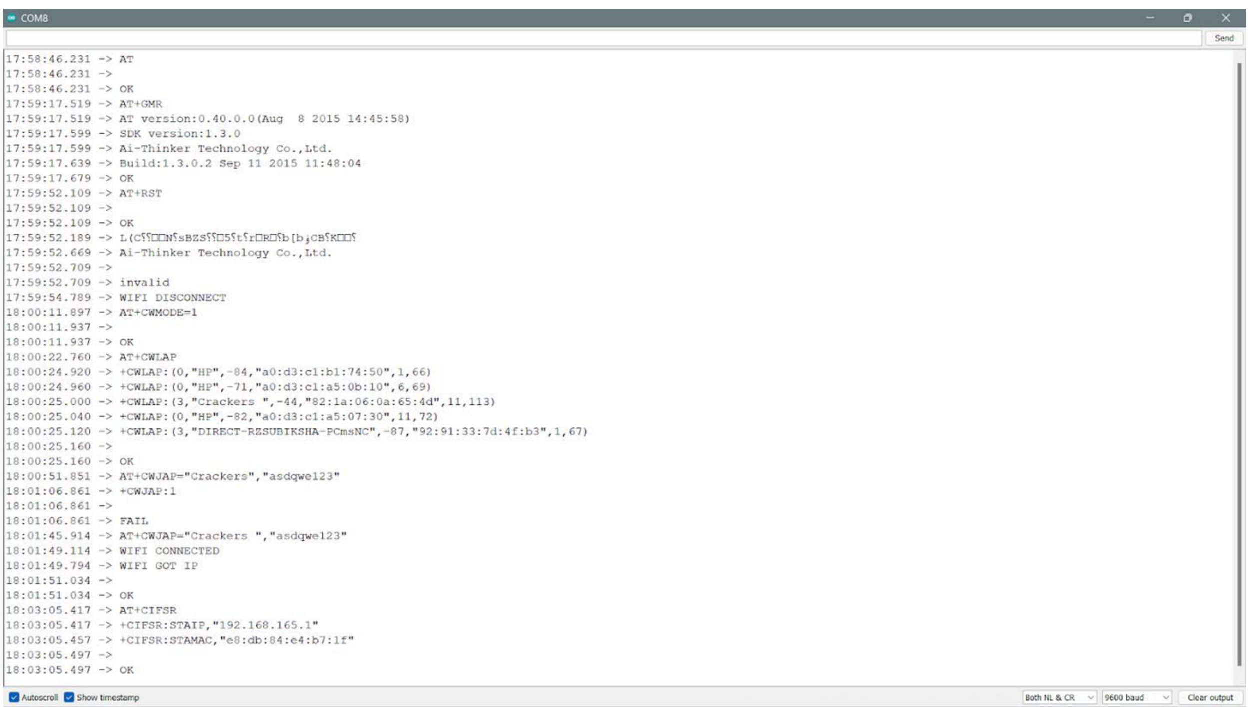

Figure 4.11 : COM window

Installing ESP8266 Into Arduino IDE And, Python Communication:

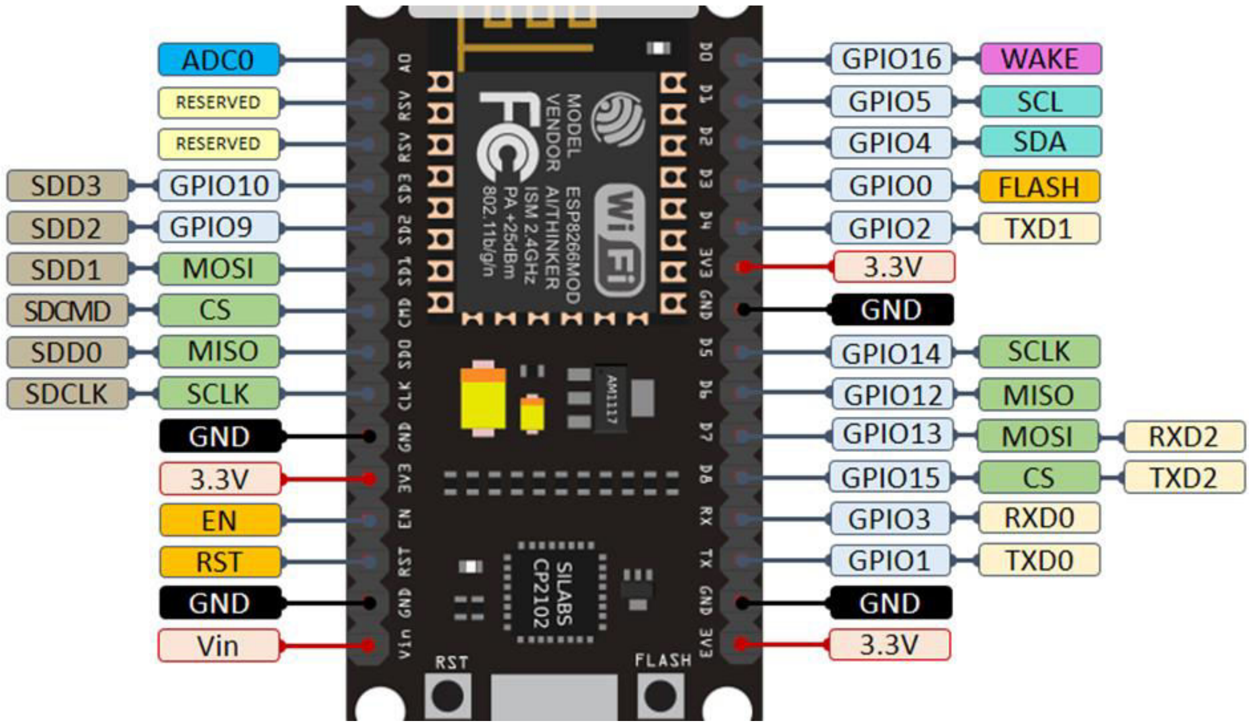

Figure 4.12 : ESP8266 12-E NodeMCU Pinout Diagram

Created a simple mDNS communication system. Our esp connects to our wifi and creates a localhost server and starts to wait for a request. Every time our python sends a request to that localhost, esp runs the desired code and then returns the result as an http request. Finally, python reads that returned data as http request and grabs those variables from it. With this, esp can return strings, datas and arrays. Python code will understand their datatype.

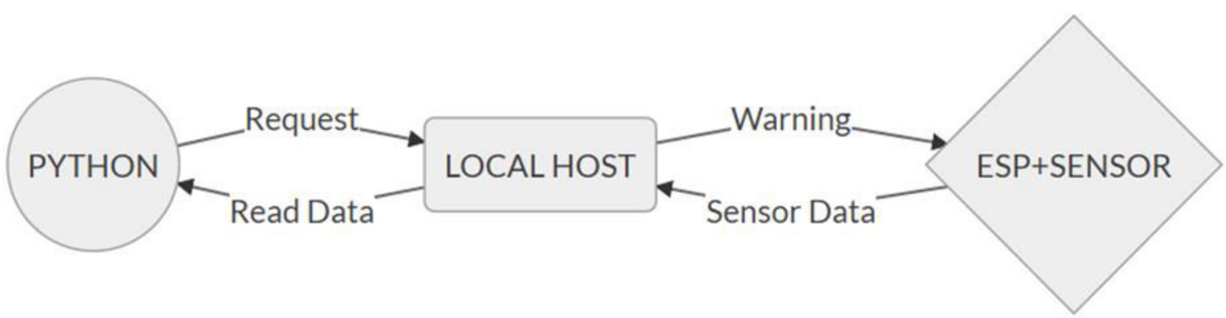

Figure 4.13 : Communication flow

While code is being uploaded, we can open the serial port, the details are printed when uploading is done. Learn the IP of esp and note that. ESP's IP on local; changes by wifi to wifi, not session to session, so when we close and open it later, it will not be changed.

We can also view the raw data with a browser while pasting the ip of ESP on a browser. Or we can make an application to read it, or can even use another language.

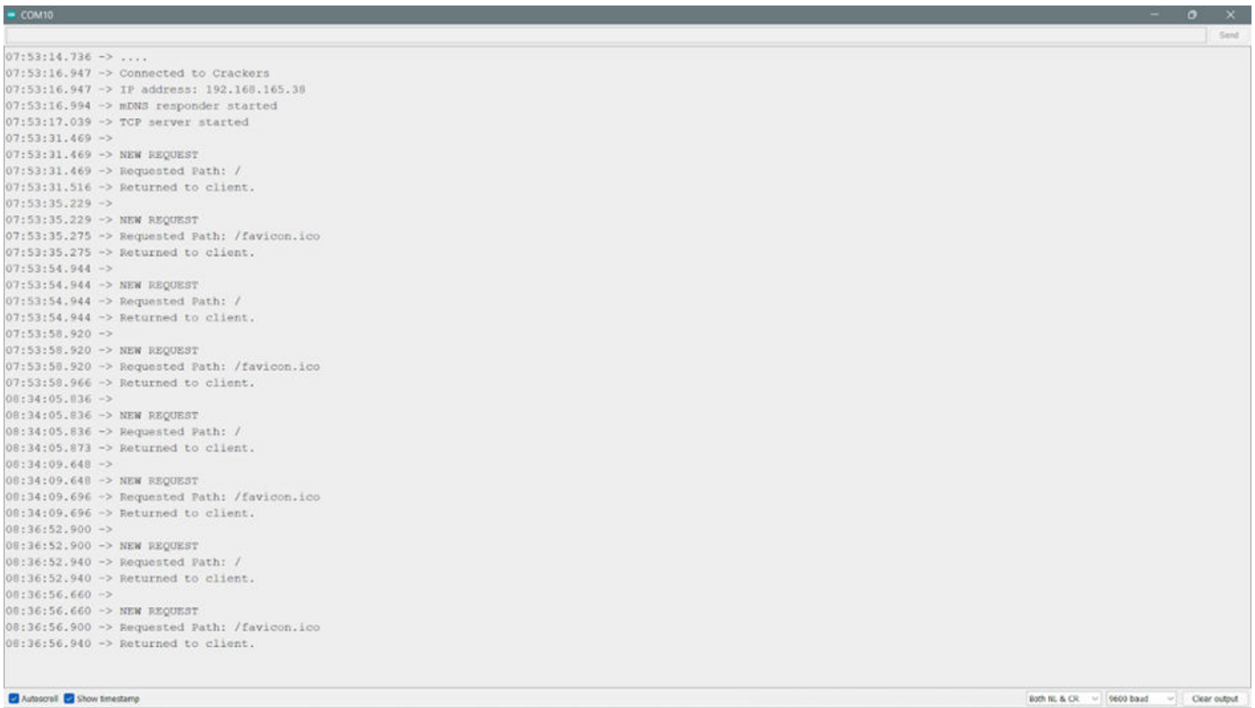

Figure 4.14 : Serial monitor window

Figure 4.15 : Browser window

4.2 Measure of SpO2, Heart Rate and BP Trends (BPT)

High physiological stress and high altitudes can be a reason for varying levels of oxygen. The human body is generally capable of adapting itself to such extreme conditions but hypoxemia is always a possibility.

This is a unique and easy way to measure PPG and calculate spO2, heart rate and Blood Pressure trending (BPT) with surprisingly high accuracy.

There are several Pulse boards available with different form factors and applications. ProtoCentral also provides the popular MAX30102-based breakout board in a variety of versions, but Pulse Express stands out by integrating the MAX32664D Biometric sensor hub.

Typically, PPG is the output from the optical sensors and it is up to the user to calculate other vitals such as Heart Rate, spO2 etc, but the MAX32664D sensor hub does all the vitals calculations and gives us the final output.

- Built in the shape and size of a finger, it is ideally suited to measuring vitals with a Velcro strap hole to connect the finger on board.
- Internal algorithm for measuring Pulse Heart Rate, Pulse Blood Oxygen Saturation (SpO2), Estimated Blood Pressure trending.
- Integrates a high-sensitivity pulse oximeter and heart rate sensor (MAX30102) and biometric sensor hub (MAX32664D).
- In-built accelerometer for robust detection and compensation of motion artefacts.

Pulse oximetry (percentage of SpO2 concentration in blood) has been used as a key health indicator for many decades. Although the original academic development was made in 1935, the modern basis for determining the SpO2 concentration using light sources and photosensor(s) was developed by Takuo Aoyagi and Michio Kishi in 1972.

When commercially feasible, SpO2 concentration measurement devices have made huge gains in medical applications. Since 1987, the Standard of Care (SoC) for the administration of a general anaesthetic has included pulse oximetry. All modern hospital bedside equipment include an SpO2 module based on the same fundamentals, albeit with minor modifications.

Pulse oximetry is used to measure the level of oxygen (oxygen saturation) in the blood. It's a simple, painless measure of how well oxygen is being sent out from our heart to parts of our body, such as our arms and legs. It can be used to check if there is enough oxygen in the blood and to check the health of a person with any condition that affects blood oxygen levels.

Blood Pressure Trend is NOT the same as absolute blood pressure measurement. BPT uses an algorithm to look at changes in the shape of the PPG signal and correlate them to changes in BP from a given calibrated baseline BP. This is still useful because it is not possible to take continuous BP recordings from traditional cuff-based sphygmomanometer devices.

Changes in BP or the BP trend (BPT) prove to be valuable in cardiovascular care and monitoring of high-risk patients.

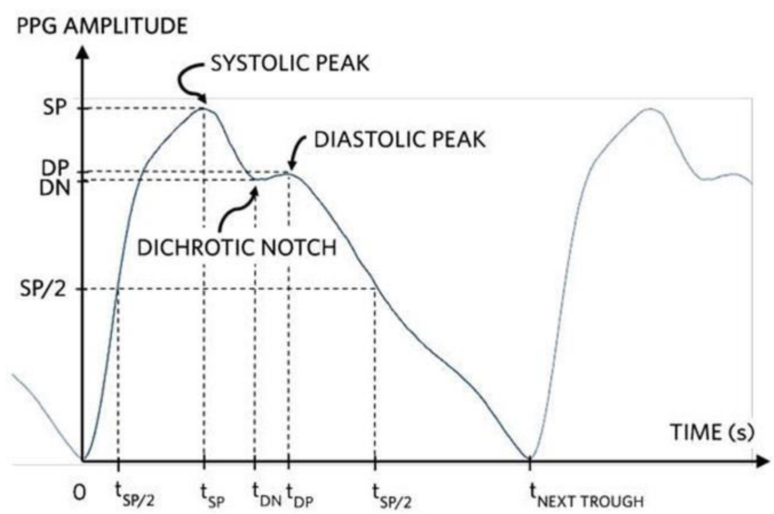

Figure 4.16 : PPG graph

Protocentral Pulse Express comes with all the basic things for us to get started with reading PPG signals and get basic measurements. It consists of a MAX32664-D Biometric Sensor Hub along with a MAX30102 pulse sensor and logic level converters. We simply have to hook it up with Arduino as shown in the following table.

We have designed the ProtoCentral Pulse Express with integrated high-sensitivity optical sensors (MAX30102) and MAX32664D biometric sensor hub to read PPG signals and perform finger based heart rate, spo2 and blood oxygen saturation measurements.

The board is connected to the MC using a standard I2C interface. Using the ProtoCentral Pulse Express Arduino Library uploaded to the NANO interfaced with the Pulse Express Pulse Oximeter and the Heart Rate Monitor makes it much easier to read the PPG and the measured SPO2, the heart rate and the estimated BP. The output is seen on the serial monitor and the serial plotter.

This code configures the sensor in algorithm mode to enable the sensor to start calculating the Spo2, BP systolic and diastolic and HR values.

The board will start in calibration mode once we upload the library, we will have to keep our finger on the sensor until the calibration progress reaches 100%. The board will switch to estimation mode once the calibration process is completed. It takes 10 to 20 seconds for the algorithm to gather enough samples to give good values.

Sensors :

- Photoplethysmography (PPG) is a simple and low-cost optical technique that can be used to detect blood volume changes in the microvascular bed of tissue. (non-invasively measurements at the skin surface)

  Pulse Sensor Module -> blood pressure and heart rate
- Signal 2mS raw Pulse Sensor signal
- DS18B20 Sensor -> Temperature
- resolution of the temperature sensor is user-configurable to 9, 10, 11, or 12 bits, corresponding to increments of 0.5°C, 0.25°C, 0.125°C, and 0.0625°C
- MPU6050 Sensor -> Accelerometer and Gyroscope
- NEO-6M GPS Module -> location of the patient

Controller : ESP8266 12-E NodeMCU

4.3 Output of sensors

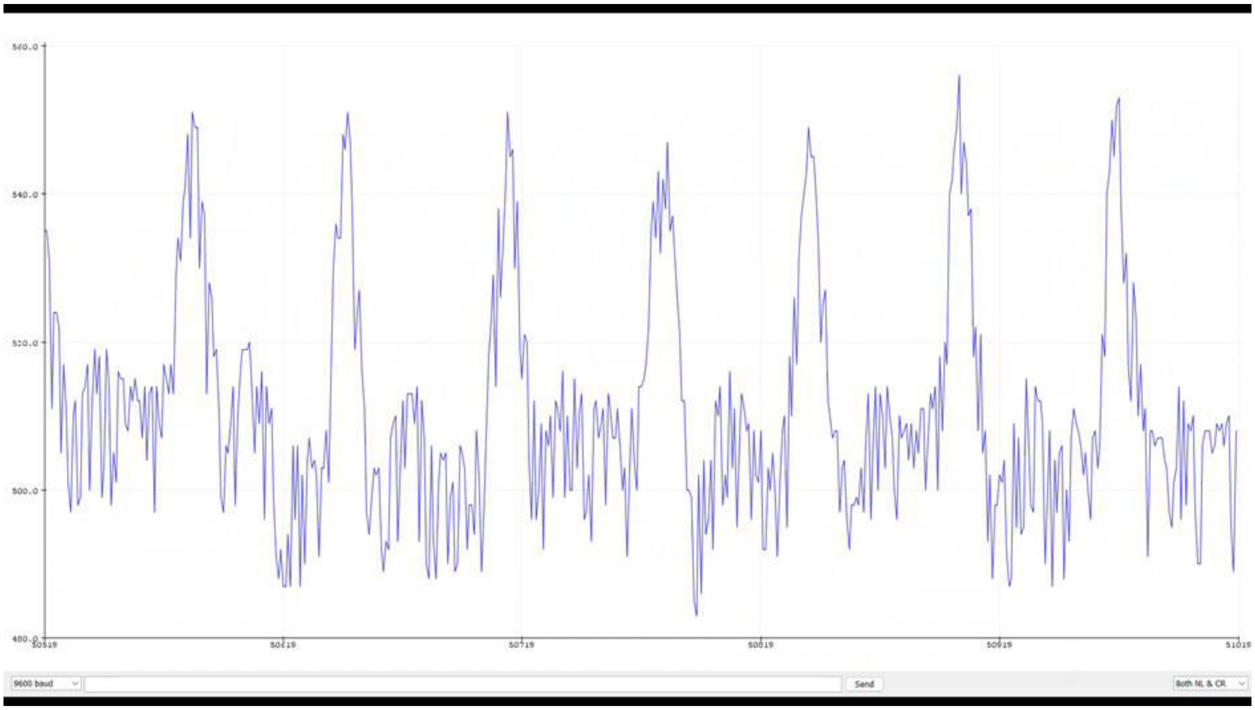

Figure 4.17 : Heart rate sensor output

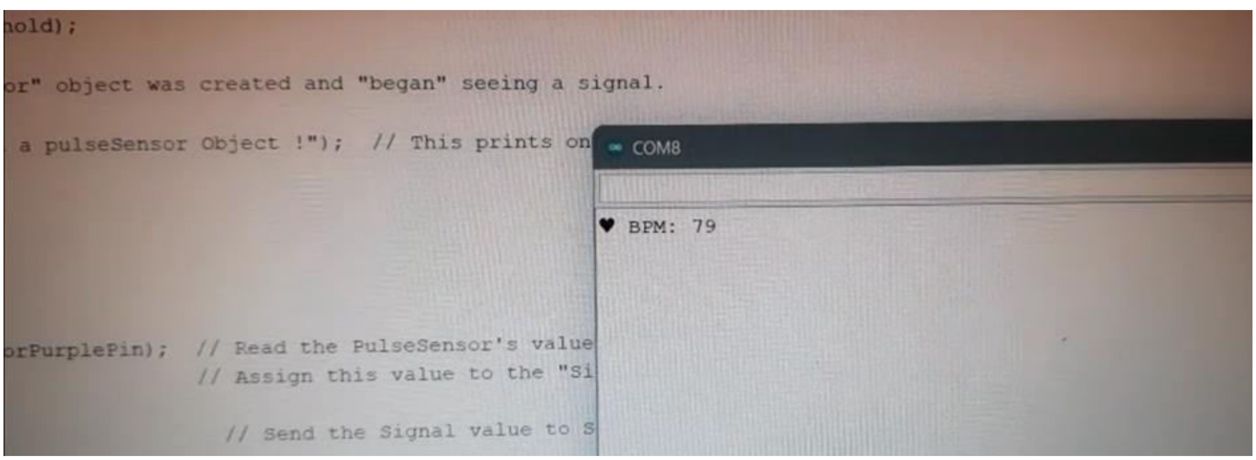

Figure 4.18 : BPM display

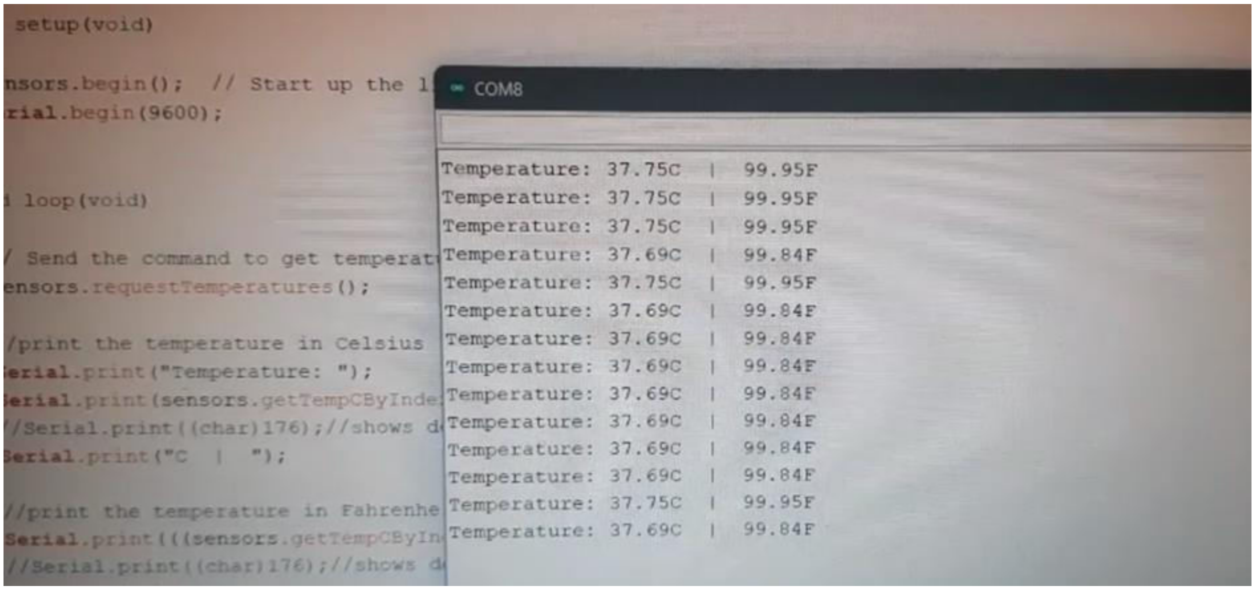

Figure 4.19 : Temperature sensor output

4.4 PCB Design

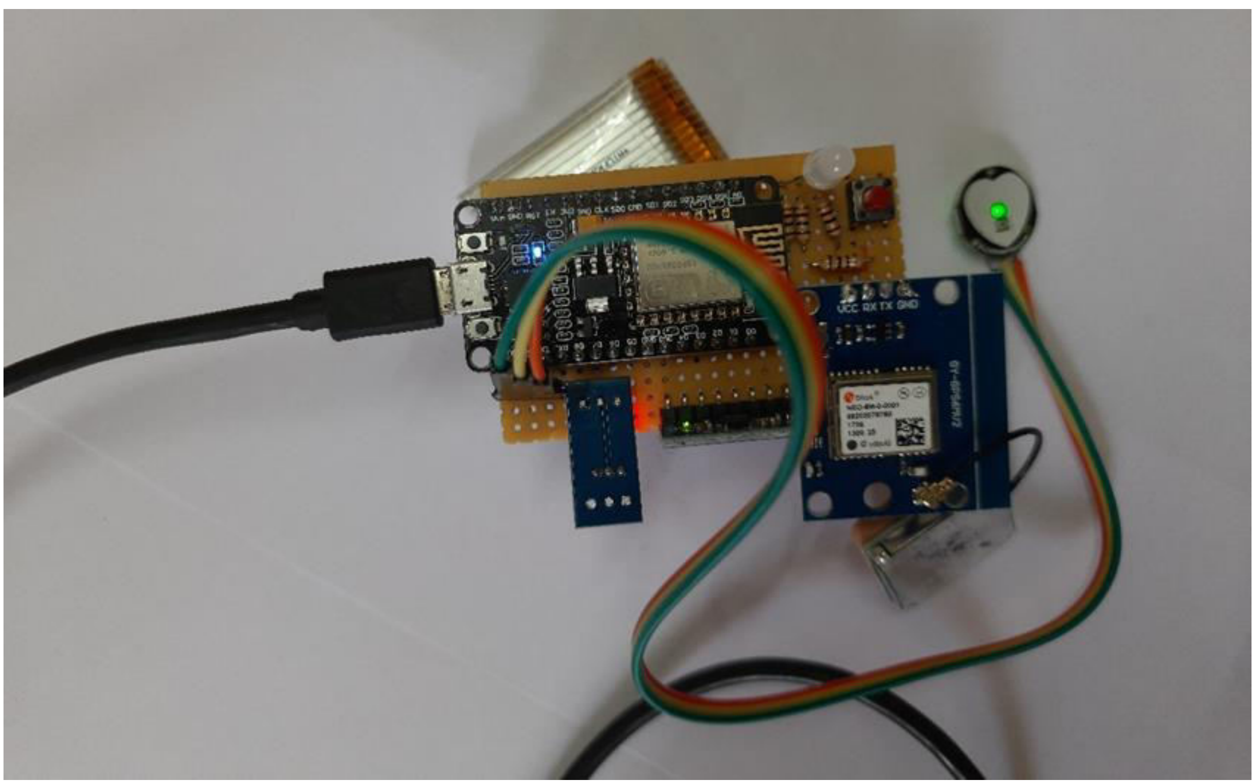

Figure 4.20 : Circuit image

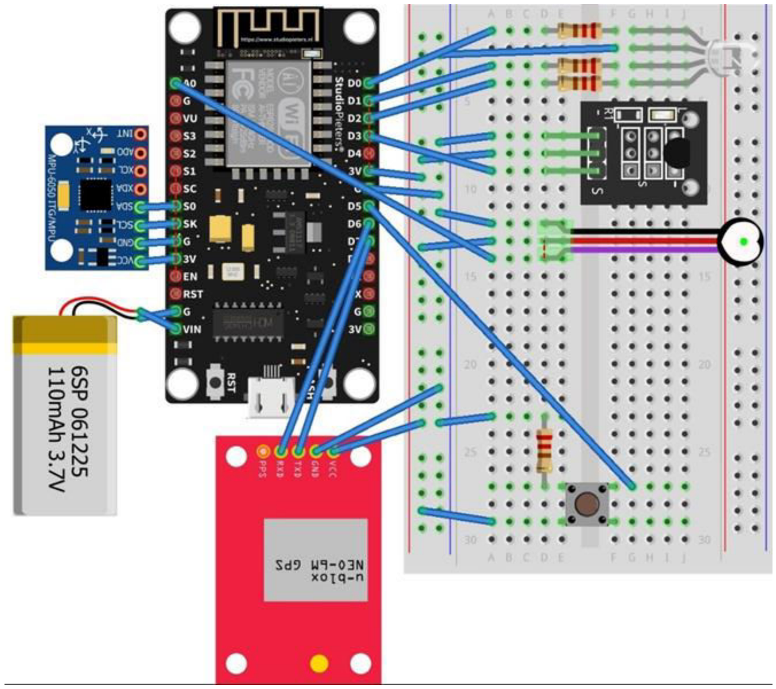

Figure 4.21 : Circuit design

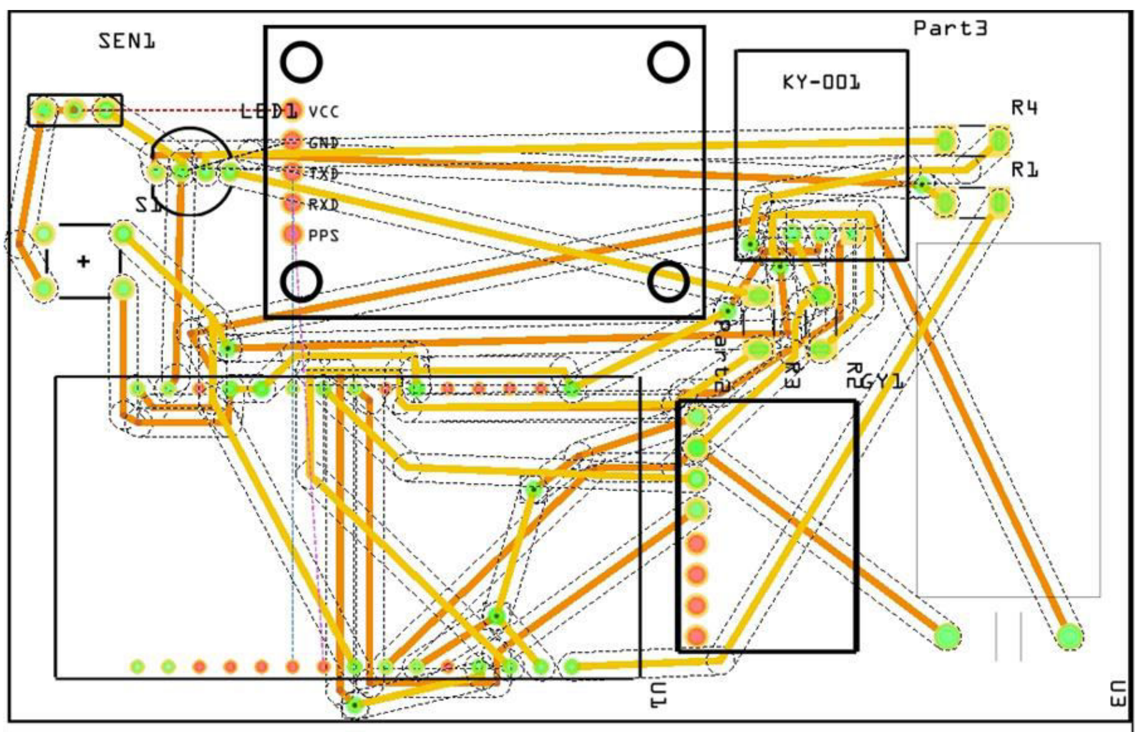

Figure 4.22 : PCB design

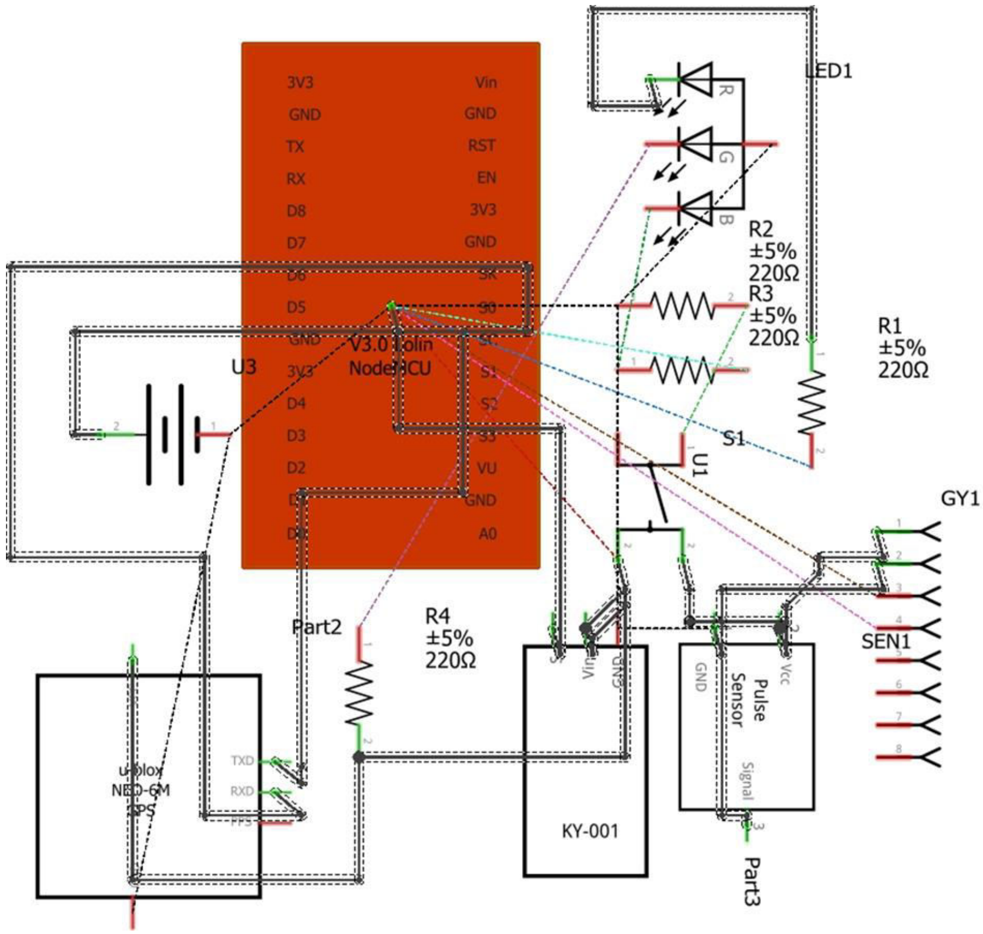

Figure 4.23 : PCB scheme

4.5 Assembly List

Label	Part Type	Properties
GY1	MPU-6050 Board GY-521	row double; form ♀ (female); package THT; variant variant 1; pins 8; hole size 1.0mm,0.508mm; pin spacing 0.1in (2.54mm)

LED1	RGB LED - (4 legs)	package 5 mm [THT]; polarity common anode; rgb RGB; pin order rgb
Part2	u-blox NEO-6M GPS Breakout	variant variant 1
Part3	KY-001 Temperature Sensor Module	chip label KY-001; editable pin labels false; variant variant 1; package Breakout board; sensor DS18B20; dimensions 18.5mm x 15mm [0.728in x 0.591in]; pins 3; temperature range -55°C to 125°C [-57°F to 257°F]; operating voltage 3.0V to 5.5V; accuracy range ±0.5°C; pin spacing 300mil
R1	220Ω Resistor	resistance 220Ω; package 2010 [SMD]; tolerance ±5%
R2	220Ω Resistor	resistance 220Ω; package 2010 [SMD]; tolerance ±5%
R3	220Ω Resistor	resistance 220Ω; package 2010 [SMD]; tolerance ±5%
R4	220Ω Resistor	resistance 220Ω; package 2010 [SMD]; tolerance ±5%
S1	Pushbutton	package [THT]
SEN1	Pulse Sensor	variant variant 1; package SIP-3
U1	NodeMCU V3.0	variant variant 1; chip ESP8266
U3	LIPO-110mAh	variant 110mAh; package lipo-110

CHAPTER 5

RESULTS

In this project we have successfully designed and tested a wearable sensor glove used to monitor the vitals of the patient wirelessly from any location. If any unforeseen situation may come up the sensor data and the location of the patient is sent to the emergency contact of the patient or to any nearby hospital which helps timely rescuing of the patient. Through the sensor data and the person's basic health records the machine learning model will predict if the person is at risk of a heart disease or not with significant accuracy.

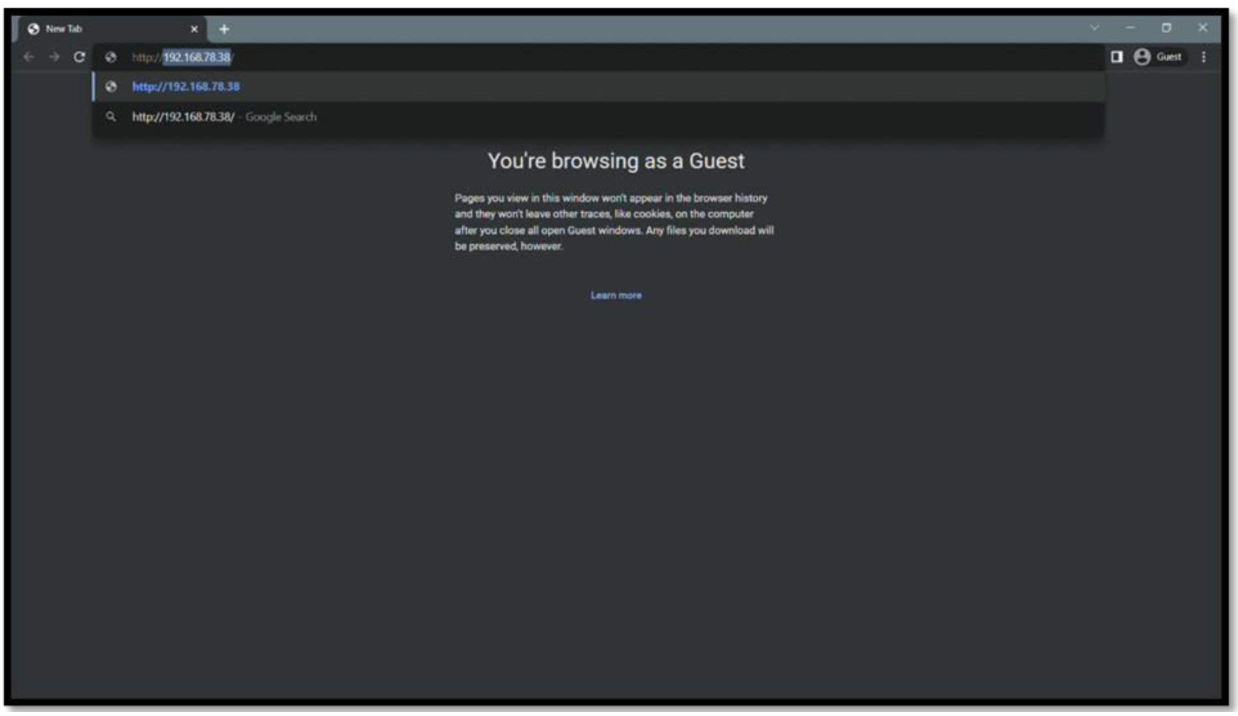

Figure 5.1 : Browser window

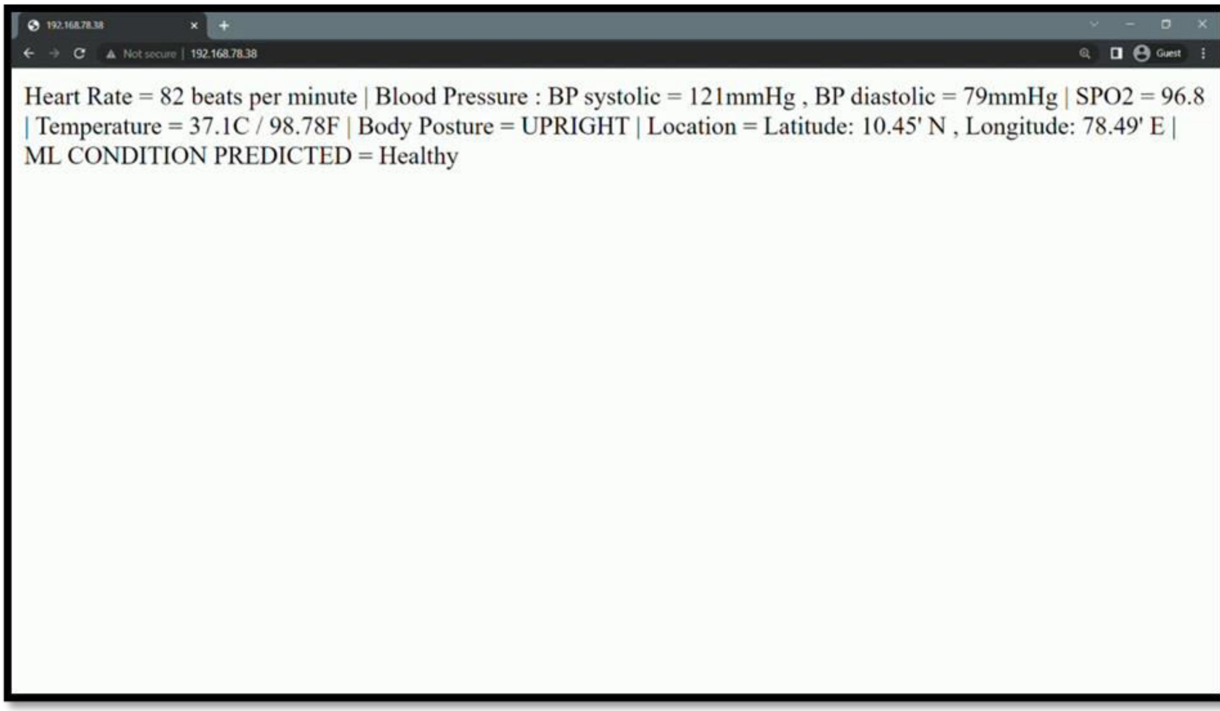

Figure 5.2 : Display of sensor values and predictions

Figure 5.3 : Display of predictions in mobile device

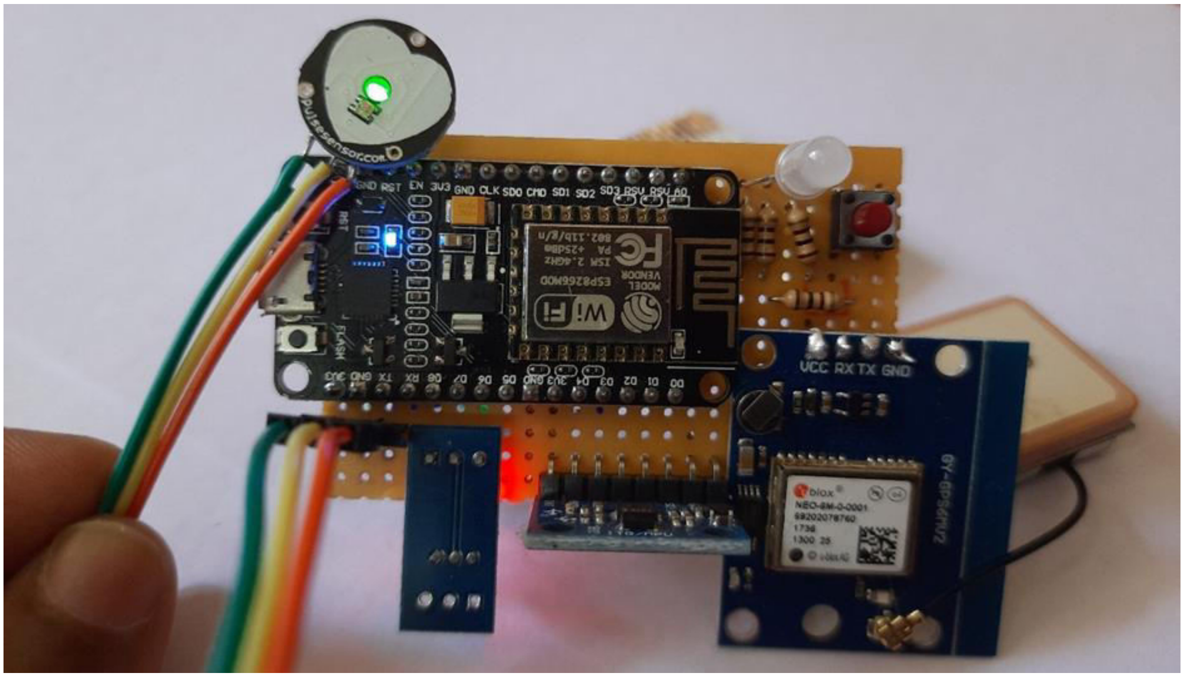

Figure 5.4 : Hardware setup

Machine learning model accuracy values:

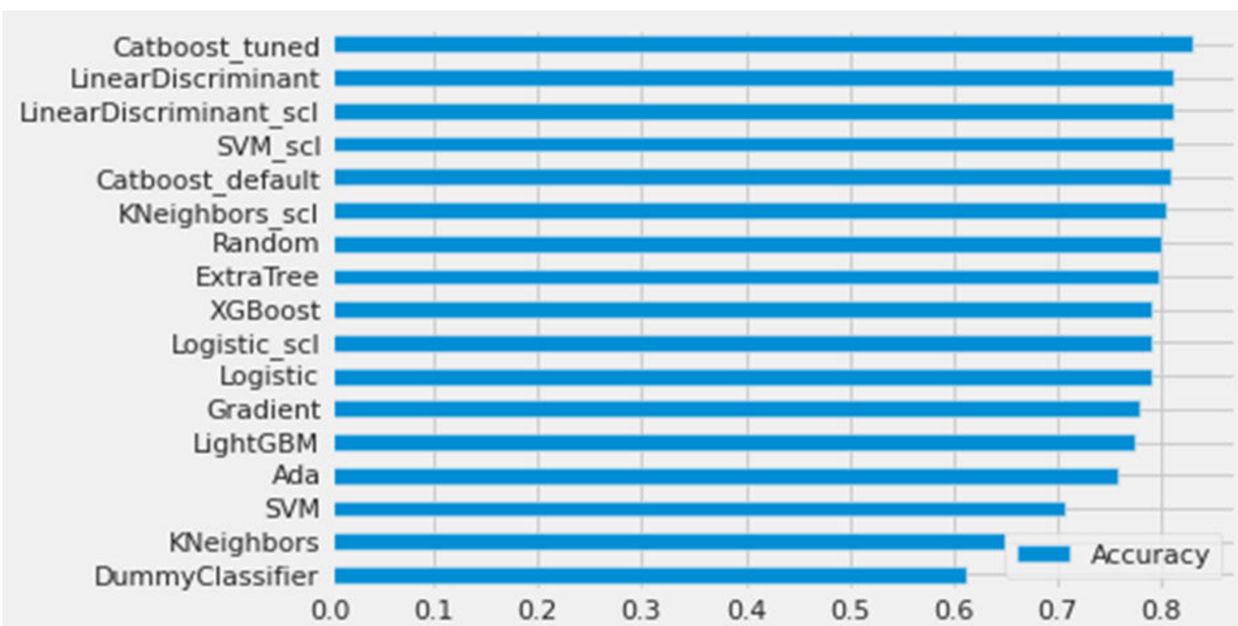

Figure 5.5 : Accuracy values

CHAPTER 6

CONCLUSION

6.1 Summary and Conclusion

In this project, an effective IoT based disease diagnosis model has been developed to monitor, predict and diagnose heart disease. An efficient framework is utilized for heart disease is created utilising the UCI Repository dataset and the healthcare sensors to predict the people who suffer from heart disease. Moreover, classification algorithms are used to classify the patient data for the identification of heart disease. Initially, the classification algorithm executes the training process which utilises the heart disease dataset to train the classifier to identify the presence of heart disease or not. Then, the trained classifier is ready to test the incoming patient details to properly identify whether the patient suffers from heart disease.

The healthcare sector is one of the most complex in terms of the level of responsibility and strict regulations, which makes it an important and vital sector for innovations. The Internet of things (IoT) has opened up a world of possibilities in the healthcare sector and could be the solution to many problems. Applying the medical IoT will bring about great opportunities for telemedicine, remote monitoring of patients' condition, and much more. This could be possible with the help of ML models. In this project, we used the most powerful algorithms, and analysed IoT and machine learning in the healthcare system to predict future risk of the patient.

6.2 Future Scope

- In the future, we will perform more experiments to increase the performance of these predictive classifiers for heart disease diagnosis by using other feature selection algorithms and optimization techniques and with the addition of more data points the model accuracy improves.

- Flexible PCB design for compact set up.
- The diversity of resources will provide better performance in knowledge extraction and a clear understanding of the measuring and collecting data problems.
- Novel methods will be designed for feature reduction to handle huge numbers of features and large volumes of healthcare records.
- A more sophisticated method will be investigated for removing irrelevant features and managing the missing values and noise to achieve efficient results.

Figure 6.1 : Device setup

REFERENCES

[1] Bo Jin, Chao Che, Zhen Liu, Shulong Zhang, Xiaomeng Yin, And Xiaopeng Wei, "Predicting the Risk of Heart Failure with EHR Sequential Data Modeling", IEEE Access 2018.

[2] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", "International Conference on "Computational Intelligence and Communication Technology" (CICT 2018).

[3] Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor, Redhwan Nour4, Samad Wali and Abdul Basit, "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection", IEEE Access 2017.

[4] Senthilkumar Mohan, Chandrasegar Thirumalai, and Gautam Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", IEEE Access 2019.

[5] K.Prasanna Lakshmi, Dr. C.R.K.Reddy, "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining", IEEE International Conference on Computer, Communication and Control (IC4-2015).

[6] M.Satish, D Sridhar, "Prediction of Heart Disease in Data Mining Technique", International Journal of Computer Trends & Technology (IJCTT), 2015.

[7] Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik, "An Intelligent Decision Support System for Cardiac Disease Detection", IJCTA, International Press 2015.

[8] Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 Vol. 2 Issue 2, February–2015.

[9] Mamatha Alex P and Shaicy P Shaji, "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique", International Conference on Communication and Signal Processing 2019.