

**EARLY DETECTION OF CARDIAC DISEASE USING MACHINE**

**LEARNING**

**BE Project Report**

**Submitted in partial fulfilment of the requirements of the Degree of Bachelor of Engineering in Information Technology by**

**Paras Chavda**

**Harsh Bhavsar**

**Yash Pithadia**

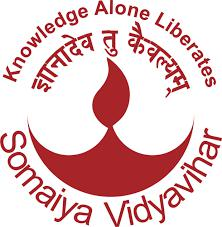
under the guidance of

**Dr. Radhika Kotecha**



**Department of Information Technology**

**K.J Somaiya Institute of Engineering and Information Technology Ayurvihar, Sion, Mumbai -400022 2017**



**CERTIFICATE**

*This is to certify that the project entitled “***EARLY DETECTION OF CARDIAC DISEASE USING MACHINE LEARNING***” is a bonafide work of Students Paras Chavda , Harsh Bhavsar and Yash Pithadia submitted to University of Mumbai in partial fulfilment of the requirement for the award of the degree of Bachelor of Engineering in Information Technology.*

| *\_\_\_\_\_\_\_\_\_\_\_\_* | *\_\_\_\_\_\_\_\_\_\_\_\_\_* |
| --- | --- |
| *Dr. Radhika Kotecha* | *Dr. Suresh Ukarande* |
| *I/c Head of Department* | *Principal* |

*Dept. of Information Technology*

Place: Sion, Mumbai-400022

Date:

**PROJECT APPROVAL FOR B.E.**

This Project report entitled “**EARLY DETECTION OF CARDIAC DISEASE USING MACHINE LEARNING**” by

Paras Chavda \_\_\_\_\_\_\_\_\_

Harsh Bhavsar \_\_\_\_\_\_\_\_\_

Yash Pithadia \_\_\_\_\_\_\_\_\_

is approved for the degree of **Bachelors of Engineering in Information Technology**.

*Examiners:*

1. *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

2. *\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

Date:

Place:

**Declaration**

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Paras Chavda\_\_\_\_\_\_\_\_\_\_\_\_\_

Harsh Bhavsar\_\_\_\_\_\_\_\_\_\_\_\_

Yash Pithadia\_\_\_\_\_\_\_\_\_\_\_\_\_

Date:

**ACKNOWLEDGEMENT**

We have taken efforts in this project. However, it would not have been possible without kind support and help of many individual and organizations. I would like to extend my sincere thanks to all of them.

We are highly indebted to Project Guide, **Dr. Radhika Kotecha** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We are thankful to our project coordinator, **Prof. Harsh Bhor** for his support.

We express our thanks to Principal, **Dr. Suresh Ukarande** for extending his support.

We would like to express my gratitude towards my parents & member of K.J. Somaiya Institute of Engineering & Information Technology for their kind cooperation and encouragement which help me in completion of this project.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

**Paras Chavda-12\_\_\_\_\_\_\_\_\_\_\_\_**

**Harsh Bhavsar-08\_\_\_\_\_\_\_\_\_\_\_\_**

**Yash Pithadia-47\_\_\_\_\_\_\_\_\_\_\_\_**

Date:

**ABSTRACT**

Heart disease has created a lot of serious concern among researches; one of the major challenges in heart disease is correct detection and finding presence of it inside a human. Early techniques have not been so much efficient in finding it even medical professor are not so much efficient enough in predicating the heart disease. There are various medical instruments available in the market for predicting heart disease there are two major problems in them, the first one is that they are very much expensive and second one is that they are not efficiently able to calculate the chance of heart disease in human. According to latest survey conducted by WHO, the medical professional able to correctly predicted only 67% of heart disease so there is a vast scope of research in area of predicating heart disease in human. With advancement in computer science has brought vast opportunities in different areas, medical science is one of the fields where the instrument of computer science can be used. In application areas of computer science varies from metrology to ocean engineering. Medical science also used some of the major available tools in computer science; in last decade artificial intelligence has gained its moment because of advancement in computation power. Machine Learning is one such tool which is widely utilized in different domains because it doesn’t require different algorithm for different dataset. Reprogrammable capacities of machine learning bring a lot of strength and opens new doors of opportunities for area like medical science. In medical science heart disease is one of the major challenges; because a lot of parameters and technicality is involve for accurately predicting this disease. Machine learning could be a better choice for achieving high accuracy for predicting not only heart disease but also another diseases because this vary tool utilizes feature vector and its various data types under various condition for predicating the heart disease algorithms such as Naive Bayes, Decision Tree, are used to predicate risk of heart diseases each algorithm has its specialty such as Naive Bayes used probability for predicting heart disease, whereas decision tree is used to provide classified report for the heart disease. All these techniques are using old patient record for getting prediction about new patient. This prediction system for heart disease helps doctors to predict heart disease in the early stage of disease resulting in saving millions of life.

|  | **FIGURE NAME** | **PAGE NO** |
| --- | --- | --- |
| 1.1 | Types of heart disease |  |
| 3.6.1 | Database Overview |  |
| 3.6.2 | Table structure of medical values |  |
| 3.6.3 | Table structure of registration details |  |
| 3.7.1 | Gantt chart representation |  |
| 3.7.2 | Project Management |  |
| 4.1.0 | Modular view of proposed system |  |
| 4.1.1 | Proposed system |  |
| 4.2 | Use case diagram |  |
| 4.3 | Activity Diagram |  |
| 4.4 | Entity Relationship Diagram |  |
| 5.1.0 | Implementation Details |  |
| 5.1.1 | Max30100 sensor |  |
| 5.1.2 | System block Diagram (max30100) |  |
| 5.1.3 | Arduino UNO |  |
| 5.1.4 | Arduino UNO interface with MAX30100 |  |
| 5.2.1 | Actual IOT implementation |  |
| 5.2.2 | Output obtained from serial monitor |  |
| 5.2.3 | Decision tree accuracy |  |
| 5.2.4 | Decision tree |  |
| 5.2.5 | Oxywatch measurements |  |
| 5.2.6 | User Interface |  |

|  | **TABLE NAME** | **PAGE NO** |
| --- | --- | --- |
|  | Dataset parameters |  |
|  | Product cost and expenses |  |

|  | **TABLE OF CONTENTS** |  |  |
| --- | --- | --- | --- |
| **Chapter No** | **Title** | **Page no.** | |
|  | List of Abbreviations |  | i |
|  | List of Figures |  | ii |
|  | List of Tables |  | iii |
| **1** | **INTRODUCTION** |  | **1** |
| 1.1 | Motivation |  | 1 |
| 1.2 | Scope |  | 3 |
| 1.3 | Aim |  | 4 |
| 1.4 | Objective |  | 4 |
| 1.5 | Significant Contribution |  |  |
| 1.6 | Organization of report |  |  |
| 1.7 | Summary |  | 4 |
| **2** | **RELATED WORK** |  | **6** |
| **3** | **ANALYSIS AND PLANNING** | | **10** |
| 3.1 | Feasibility Study | | 10 |
| 3.2 | Principal actors |  | 13 |
| 3.3 | User characteristics |  | 14 |
| 3.4 | Assumptions and Dependencies |  | 16 |

| 3.5 | Requirement Analysis | 17 |
| --- | --- | --- |
| 3.5.1 | Functional Requirements | 18 |
| 3.5.2 | Non-Functional Requirements | 19 |
| 3.5.3 | Interface Requirements | 20 |
| 3.6 | Database Design | 21 |
| 3.7 | Project Scheduling |  |
| **4** | **PROPOSED WORK** | **22** |
| 4.1 | Proposed Approach | 22 |
| 4.2 | Use Case Diagram | 22 |
| 4.3 | Activity Diagram |  |
| 4.4 | Entity Relationship Diagram |  |
| 4.5 | Product Cost and Expenses |  |
| **5** | **IMPLEMENTATION DETAILS AND RESULTS** | **23** |
| 5.1 | Implementation details | 23 |
| 5.1.1 | Hardware implementation | 24 |
| 5.1.1.1 | Data Acquistation Unit | 25 |
| 5.1.1.2 | Microcontroller Unit |  |
| 5.1.1.3 | Display Unit |  |
| 5.1.2 | Software design |  |
| 5.1.2.1 | Arduino IDE | 25 |
| 5.1.2.2 | Python libraries | 25 |
| 5.1.2.3 | Jupyter Notebook | 25 |
| 5.1.2.4 | Apache Cordova | 25 |
| 5.1.2.5 | Flask | 26 |
| 5.2 | Results and Discussions | 26 |
| **6** | **CONCLUSION AND FUTURE WORK** | 27 |
| **7** | **REFERENCES** | 27 |

**CHAPTER 1**

**INTRODUCTION**

Heart is the most important organ in living beings; if it does not function properly it will affect other parts of body. As per the estimation of World Health Organization (WHO) 12 million deaths occur due to heart diseases, and nearly 80% of death in the world will be because of heart disease .Now a days at many places clinical test results are often based on doctor’s intuition and experiences rather than information available in many large databases. Hence this process leads to unintentional biases, errors and a huge medical cost which affects the quality of service provided to patients. The term "heart disease" is often used interchangeably with the term "cardiovascular disease." Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease. With the increasing number of population across the world and with recent changes of humans' life styles, there is increasingly higher numbers of individuals with complex medical conditions. This has led to higher numbers of people visiting hospitals and put stress on the Medicare health systems. Thus, there is an increasingly need for health care systems that can assist with these challenges.

Today many hospitals have installed databases systems to manage their clinical data or patient data. These information systems typically generate large amounts of data which can be in any format like numbers, text, charts and images but unfortunately, this database that contains rich information is rarely used for clinical decision making. There is much information stored in repositories that can be used effectively to support decision making in healthcare. Data mining techniques is widely used in medical field for extracting data from database. In data mining decision tree is a method which is used extensively. Decision trees are non-parametric supervised learning method used for classification. The main aim is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The structure of the decision tree is in the form of tree and leaf nodes. Decision trees are most commonly used in operations research, mainly in decision analysis. Advantages are that they are easy to understand and interpret. They are robust, performed well with large datasets, able to handle both numerical and categorical data. By providing efficient treatments, it can help to reduce costs of treatment. Using data mining techniques it takes less time for the prediction of the disease with more accuracy.

Recently, there has been growing attention to the advances in the areas of electronic and biomedical engineering and the great applications that these technologies can offer mainly for health diagnosis monitoring & analysis. Wearable sensors are now accessible for many people worldwide with affordable prices. These devices in conjunction with artificial intelligence techniques can be effective for prediction & diagnosis of people with heart diseases with improving people’s lives.

**The Risk Factors of Heart Disease are:**

**1)Heredity:**Most people know that the heat disease can run in families. That if anybody has a family history of heart disease, he/she may be at greater risk for heart diseases.

**2)Smoking:** Smoking is major cause of heart attack. Nearly 40% of all people who die from smoking tobacco do so due of heart and blood vessel diseases. A smoker’s risk of heart attack reduces rapidly after only one year of not smoking. It damages the lining of arteries and builds up a fatty material called atheroma which narrows the arteries causing heart attacks.

**3)Cholesterol**: Cholesterol is a soft, waxy substance found among the lipids in the bloodstream and in all the body’s cells, abnormal levels of lipids (fats) in the blood are risk factor of heart diseases. High level of triglyceride (most common type of fat in body) combined with high levels of LDL (low density lipoprotein) cholesterol speed up atherosclerosis increasing the risk of heart diseases. Increase in the fatty deposits (high cholesterol) does not allow sufficient blood to flow through the arteries thus causing heart attacks.

**4)High Blood Pressure:** when the heart pumps blood, the force of the blood pushes against the walls of the arteries causing pressure. If the pressure rises and stays high over the time it is called high blood pressure or hypertension. This can harm the body in a way by increasing the risk of getting heart disease.

**5)Obesity:** The term obesity is used to describe the health condition of anyone significantly above his or her ideal healthy weight. Being obese puts anybody at a higher risk for health problem such as heart disease, stroke, high blood pressure, diabetes and more. Obese people are more likely to have high blood pressure, high cholesterol level and diabetes (increase in blood sugar level) which increases the risk of heart strokes in human body.

**6)Lack of Physical Exercise**:lack of exercise is a risk factor for developing various diseases. The increase in the blood pressure and increases cholesterol level in blood vessels increases the risk of heart diseases

**Types of Heart Disease** : There are many types of heart disease that affect different parts of the organ and occur in different ways.

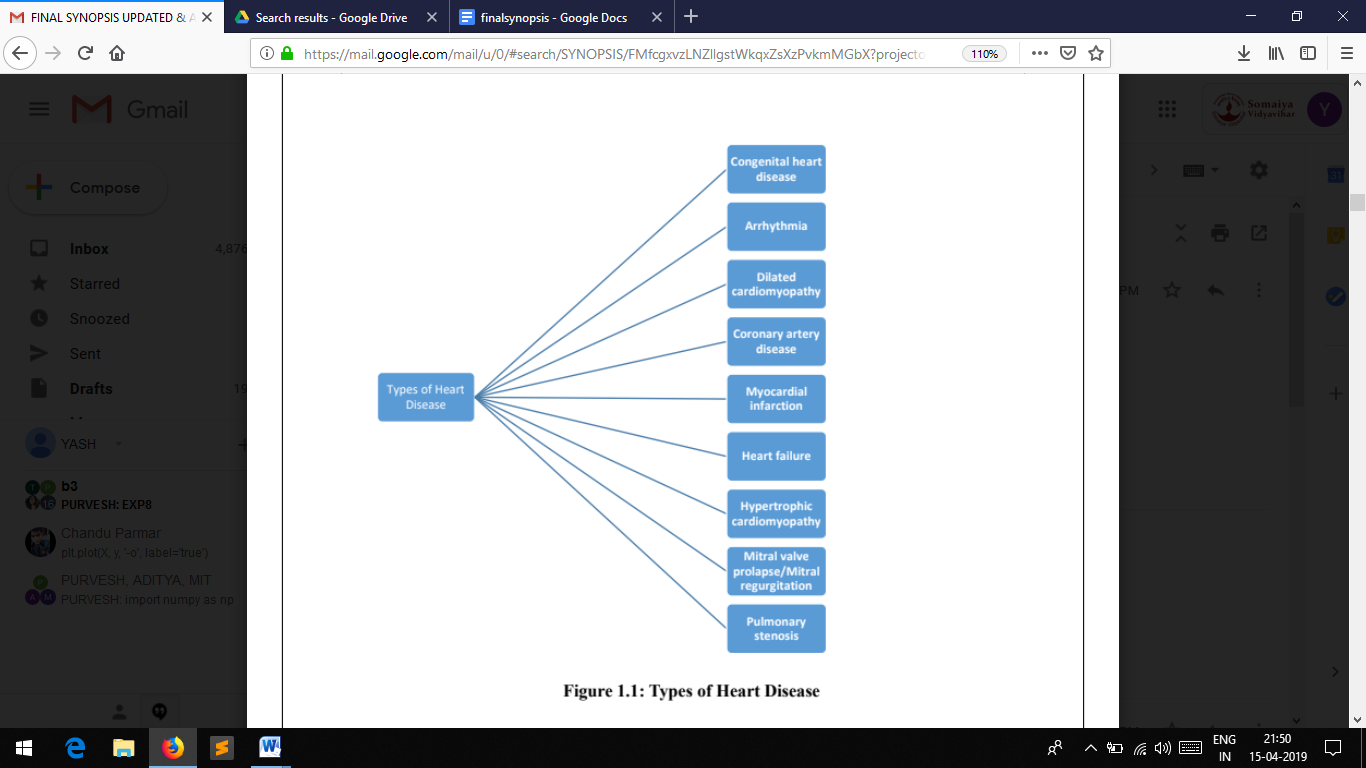


Fig. 1.1: Types of heart disease

Fig. 1.1, depicts the types of heart disease which can lead to a heart attack or a heart failure.

**Congenital heart disease**: This is a general term for some deformities of the heart that have been present since birth. Examples include: Septal defects**:** There is a hole between the two chambers of the heart. Obstruction defects: The flow of blood through various chambers of the heart is partially or totally blocked. Cyanotic heart disease: A defect in the heart causes a shortage of oxygen around the body.

**Arrhythmia:** Arrhythmia is an irregular heartbeat. There are several ways in which a heartbeat can lose its regular rhythm. These include: tachycardia, when the heart beats too fast bradycardia, when the heart beats too slowly premature ventricular contractions, or additional, abnormal beats fibrillation, when the heartbeat is irregular. Arrhythmias occur when the electrical impulses in the heart that coordinate the heartbeat do not work properly. These make the heart beat in a way it should not, whether that be too fast, too slowly, or too erratically. Irregular heartbeats are common, and all people experience them. They feel like a fluttering or a racing heart. However, when they change too much or occur because of a damaged or weak heart, they need to be taken more seriously and treated. Arrhythmias can become fatal.

**Coronary artery disease:** The coronary arteries supply the heart muscle with nutrients and oxygen by circulating blood. Coronary arteries can become diseased or damaged, usually because of plaque deposits that contain cholesterol. Plaque build-up narrows the coronary arteries, and this causes the heart to receive less oxygen and nutrients.

**Dilated cardiomyopathy:** The heart chambers become dilated as a result of heart muscle weakness and cannot pump blood properly. The most common reason is that not enough oxygen reaches the heart muscle, due to coronary artery disease. This usually affects the left ventricle.

**Myocardial infarction:** This is also known as a heart attack, cardiac infarction, and coronary thrombosis. An interrupted blood flow damages or destroys part of the heart muscle. This is usually caused by a blood clot that develops in one of the coronary arteries and can also occur if an artery suddenly narrows or spasms.

**Heart failure:** Also known as congestive heart failure, heart failure occurs when the heart does not pump blood around the body efficiently. The left or right side of the heart might be affected. Rarely, both sides are. Coronary artery disease or high blood pressure can, over time, leave the heart too stiff or weak to fill and pump properly.

**Hypertrophic cardiomyopathy** : This is a genetic disorder in which the wall of the left ventricle thickens, making it harder for blood to be pumped out of the heart. This is the leading cause of sudden death in athletes. A parent with hypertrophic cardiomyopathy has a 50 percent chance of passing the disorder on to their children.

**Mitral regurgitation:** Also known as mitral valve regurgitation, mitral insufficiency, or mitral incompetence, this occurs when the mitral valve in the heart does not close tightly enough. This allows blood to flow back into the heart when it should leave. As a result, blood cannot move through the heart or the body efficiently.

**Mitral valve prolapse:** The valve between the left atrium and left ventricle does not fully close, it bulges upwards, or back into the atrium. In most people, the condition is not life-threatening, and no treatment is required. Some people, especially if the condition is marked by mitral regurgitation, may require treatment.

**Pulmonary stenosis:** It becomes hard for the heart to pump blood from the right ventricle into the pulmonary artery because the pulmonary valve is too tight. The right ventricle has to work harder to overcome the obstruction. An infant with severe stenosis can turn blue. Older children will generally have no symptoms. Treatment is needed if the pressure in the right ventricle is too high, and a balloon valvuloplasty or open-heart surgery may be performed to clear an obstruction.

**1.1 Motivation**

The motivation of this work comes from the fact that Cardiovascular Disease (CVD) is the number one cause of death globally, more people die annually from CVDs than from any other cause. Another motive to develop the proposed system was that, in the current system a person who intends to check his cardiac health needs to visit the doctor regularly and get all the medical tests done as per advise of the doctor and show the respective reports to him. Depending on the status of the reports, the doctor examines the patient and if found problematic, the patient is treated accordingly. Apart from this there is no way by which an early detection of a cardiac disease can be done. Especially in India, the number if people died due to heart disease is much more than the casualties caused in the European countries [1]. Deaths due to cardiac disease in has increased continuously in India, it rose by around 41 percent from 155.7 to 209.1 deaths per one lakh population [2]. In children, heart failure can present as respiratory distress, easy fatigability, poor tolerance to exercise, etc.

The pervasiveness of cardiac disease and stroke has rapidly increased by over 50% from 1990 to 2016 in India, with an increase observed in every state. As a result, it can be stated that the number of total deaths and disease burden in the country has almost doubled in the past 25 years [3]. The ratio of deaths and disability because of heart disease was significantly higher in men compared to that in women. There was a rapid increase in deaths due to cardiovascular diseases which was accounted to rise from 13 lakh in the year 1900 to 28 lakh in the year 2016 [3]. The World Health Organization (WHO) has estimated that, with the current burden of CVD, India would lose $237 billion from the loss of productivity and spending on health care over a 10-year period (2005–2015) [4]. Some of the major causes that stimulate the probability of cardiac disease includes: harmful use of alcohol, active smoking raised blood pressure, high cholesterol level excessive presence of fats, lack of physical exercise, high sugar level and an unhealthy diet. The amount of deaths due to cardiac disease was the highest in states like Kerala, Punjab and Tamil Nadu, followed by Andhra Pradesh, Himachal Pradesh, Maharashtra, Goa, and West Bengal.

More than half of the total deaths due to any cardiovascular disease in India in 2016 were in people younger than the age of 70 years. This proportion of death was the highest in less developed states compared with developed and underdevelopment states, which are a major cause for concern with respect to the challenges, posed to the health systems. As a result, reducing premature deaths from cardiovascular diseases in the economically productive age groups requires urgent action across all states of India.

* 1. **Scope**

The goal is to develop a low power, more reliable, non-intrusive, are the essential signs monitor which gather information on the body and send the parameters to a display module, proposing a remote monitoring and sensing parameter of the human body which consists of heart rate and Spo2. The parameters that are used for sensing and observing will send the data through sensors to oled display.

We used Decision trees as they implicitly perform feature selection & can tackle nonlinear relationships between parameters. Each leaf of the tree is labeled with a class or probability distribution over the classes. A tree can be "learned" by splitting the source set into subsets based on attributes and the recursion is completed when splitting no longer adds value to the predictions. The information gained is based on the decrease in “entropy” after dataset is split. Following equation shows the formula for entropy, where ‘p’ is the probability of certain class occurring, given a specific feature. Decision tree is used here as a classification technique, which trains the model with pre-fitted data values and with respect to the values entered by the user, it predicts the occurrence of a heart disease in an individual.

**1.3 Aim**

This project comprises the development of a framework based on decision tree classification techniques on heart dataset for early diagnosis of heart based diseases. It is hard to diagnose the heart diseases with just observation that arrives suddenly and may prove fatal when it’s uncontrolled. Proposing a remote sensing parameter of the human body which consists of pulse and temperature. The parameters that are used for sensing and monitoring will send the data through sensors.

**1.4 Objective**

Full term objective of this project is to be able to make a system which predicts an occurrence of a heart disease and notify the user what symptom has led the diagnosis of a heart disease.

The short term objectives of the project are as follows

* To fetch the real time data from the users by using Arduino.
* Train the machine using dataset obtained from UCI repository.
* Cleaning the dataset.
* Prediction of heart disease using the data obtained from the users as well as the cleaned dataset.
* Optimizing the algorithm.

**1.5 Significant Contributions**

Significant contribution from this project can be considered in a form of a system which helps the individual to maintain and monitor the cardiac health. System comprises of two modules which are the hardware module and the software module. The hardware module is used to provide user’s real time data about their heart rate and Spo2 level measured with the help of Arduino and Max30100 sensor. The software module accepts various other parameters from the user to efficiently predict the results about the occurrence of heart disease by using decision tree classifier

**1.6 Summary**

Heart related diseases are primarily the main reason of death throughout the world and due to which a large number of casualties are arising in countries with low and middle income like India. A large amount of data is continuously generated by medical practitioners. The prediction of heart disease is most complicated task in the field of medical sciences which cannot be observed with a naked eye and comes instantly anywhere, anytime. So there arises a need to develop a decision support system for detecting heart disease. A heart disease prediction model using data mining technique, called decision tree algorithm which helps medical practitioners in detecting the disease based on the patient’s clinical data. In this project, we propose an efficient decision tree algorithm technique for heart disease prediction. To achieve correct and cost effective treatment computer-based systems can be developed to make good decision. The data generated can be used for the early detection of cardiac diseases, which can effectively support to reduce the occurrence of various heart related diseases. The decision prediction can be effectively done by enhancing the knowledge identification required to discover patterns that were not formerly known. Efficient prediction can be done by accessing the data accumulated from health care companies and industries and find the hidden patterns. The proposed work uses a machine learning algorithm on cardiac-related data and attempts to detect the possibility of cardiac diseases prior to suffering from serious issues.

**CHAPTER 2**

**RELATED WORK**

In [5], the Adaptive Network based Fuzzy Inference System and Linear Discriminant Analysis (ANFIS-LDA) model combined a fuzzy inference method and LDA to predict Coronary Heart Disease (CHD), thus increasing the specificity, sensitivity, and accuracy of the model. In order to increase the prediction accuracy, 625 rules were created by using sample medical data. Further, the classification performance was improved using an ANFIS and LDA. The results of the proposed model showed higher accuracy than those of the existing models. Thus, the proposed model can be used for the prevention of CHD in the general public.

In [6], a mobile monitoring solution is proposed that addresses these challenges and incorporating some smart features to encounter the energy insufficiency of mobile devices and network interruption. The authors of the paper have developed a formal model that evaluated the best execution decision considering online, offline, or combined processing. The model used Dynamic Programming (DP) to determine the best execution path that guarantees an optimum execution time given the resources constraints mentioned above. The paper evaluated the applicability of our solution using electrocardiograph dataset, and proposed paper evaluated the key monitoring processes including preprocessing, feature extraction, and classification.

In [7], the target is to determine the aspects of use of healthcare data which can come to the aid of people by machine learning methods and data mining procedures. The primary objective is to suggest an automated system for diagnosing heart diseases by taking into account earlier information and data. To find out few vital and basic inquiries related with healthcare organization data mining methods are used. For the prediction of heart disease classification models of data mining like Decision tree, Neural networks and Naive Bayes classifier are applied.

In [8], data mining plays an inevitable role in the prediction process of many chronic diseases including deadly heart related diseases. The proposed study examined and revealed the results of applying both k Nearest Neighbor (kNN) and random forests to the Framingham scoring model designed for early risk prediction of Hard Coronary Heart Disease (HCHD). The results reflected that the accuracy of (kNN) was higher compared to random forests in identifying the risk classes among the test dataset compared to the training one. However, the accuracy rate is still to be improved. The authors of the paper suggest using the enhanced kNN approach to enhance the classifier’s performance as future scope of work.

In [9], analysis was performed using Classification And Regression Tree (CART) algorithm that yielded ~80% accuracy which was analyzed comparatively well to the performances of different classifiers like Support Vector Machine or Artificial Immune Recognition System (AIRS). From our CART graphical model, it was found that the basis of the tree was feature fifteen, or the center rate in range of beats per minute. As the root of the tree is recommended to be a determinant feature of the info, this is smart as a result of expected traditional or abnormal heart rates to be strongly correlated with arrhythmia. Looking at the top nodes of the tree, it was noticed that these nodes were related to a multiple of different features, but mainly features in the 200 range i.e. channel values.

In [10], before feature selection, Naive Bayes achieves lower cross validation error than SVM. While once feature choice, SVM achieves lower cross validation error than Naive Bayes. The proposed system thinks that the matter could belong the shortage of enough coaching examples (475) and excessive quantity of features (274). Just as it will be seen in drawback, set a pair of once each Naive mathematician and SVM are used to classify spam and non-spam, SVM formula works higher if there are more training examples. However, during this drawback, solely 475 coaching examples are accessible.

In [11], we have studied about a health-platform aimed to improve the lives of chronic-disease sufferers. Besides the improvement of the decision-making process based on the personal history of the patient, which is a common feature on monitoring systems, this medical platform assures that the decision is taken at the proper time by continuously analyzing data in real-time. The distinctive element in the current paper is the way in which the data is analyzed based on the complex criteria that the physicians can create. Future work will be focused on creating and analyzing use cases to improve the system’s functionalities. Current research presented in this paper was focused on creating a viable, reliable and secure solution of analyzing patient’s health data in real-time.

In [12], we learned that Decision Tree is one of the successful data mining techniques used in the diagnosis of heart disease. Yet its accuracy is not perfect. Most research applies the J4.8 Decision Tree that is based on Gain Ratio and binary discretization. This research systematically tested combinations of discretization, decision tree type and voting to identify a more robust, more accurate method. The supervised discretization methods do not show any enhancement in the Decision Tree accuracy either with or without voting. Applying voting shows increase in the accuracy of different types of Decision Tree. Systematic testing against a widely-used benchmark data set shows that nine voting with equal frequency discretization and Gain Ratio Decision Tree can enhance the accuracy of the diagnosis of heart disease.

In [13], we observed that The objective of the proposed research is to search the best combination of classifiers in an ensemble that is generally suitable for all data sets of Heart diseases diagnosing. Subsequently, six classifiers like Bayesian Net, Naive Bayes, Support Vector Machine, Neural Network, C4.5 and FDT are used to predict two different heart diseases. Observations reveal that the best combination for both datasets is mostly the combinations which the Naive Bayes is one of its classifiers with accuracy of 92%. In future, they aimed to reach a unified data set for most of the heart diseases instead of having different data sets with different attributes from different resources in order to globalize the combination of classifiers proposed to be used for any dataset in the domain of heart.

**CHAPTER 3**

**ANALYSIS AND PLANNING**

**3.1 Feasibility Study**

Feasibility study is an important phase in the project development process. It enables the developer to have an assessment of the product being developed. It refers to the feasibility study of the product in terms of outcomes of the product, operational use and technical support required for implementing it.

Early detection of cardiac disease using machine learning is feasible if technology and system feasibility are taken into concern. As technology is modernizing day-by-day, Modern technology is used in this project so that it is user-friendly. Our project is developed using a cross hybrid app development platform as a result it can be using by users having any operating system like Android, Apple etc. A user-friendly GUI is used at Early detection of cardiac disease using machine learning so that user feels comfortable to use.

**Economic Feasibility:**

**1.                  Development Cost**

           The development cost of Early detection of cardiac disease using machine learning is ₹1800 that includes both software and hardware cost. It includes hardware like Arduino uno, Max30100 sensor, OLED display, etc and software like Apache Cordova, Flask, Html, Css. Mostly all the hardware and software that we used are open-source and few of them are free to use. We have tried to make our development cost as low as possible to make it cost-efficient.

**2.                  Operational Cost**

    Early detection of cardiac disease using machine learning only need a connection to a power supply to power up the Arduino module which helps in measuring the heart rate as well as Spo2 level. This power supply can easily provided by using a mobile charger and the connecting cable. As a result the system can be powered with a minimum electricity cost.

**Operational Feasibility:**

Each and every technology that is developed is operationally feasible so it is easy to use. We have tried to develop “Early detection of cardiac disease using machine learning” operationally feasible so that even a non-technical person can use it without any help.

**3.5  Requirement Analysis**

It is far much important to analyze requirements before-hand  starting any project’s tasks. In current scenario every customer needs a product or service which can be used at ease and should be efficient. Since manually going to the doctor and taking reports related to heart rate and temperature, which is not a static entity, can be time consuming. So using pulse and temperature sensors we can bring advancements in the existing methodologies. That is the patient does not need to visit the clinic to check his/her heart rate. The heart rate of the patient and the temperature around the patient can be calculated using sensors. Now this entities the user can calculate himself using appropriate sensors.

For the user to access the desired output of the project an application is made. The sole purpose of it is to make the patient access the application at any point of time. Using a website is not feasible for every patient to access since not everyone has a computer but everyone carries a smartphone in today’s world. Well websites can be accessed on the smartphone. True, but the way any application provides the ease to users is far much better than any website accessed over smartphone.

**3.5.1 Functional Requirements :**

The sensors are implemented using Arduino UNO. The UNO is the most used and documented board of the whole Arduino family. Arduino Uno is a microcontroller board based on the ATmega328P ([datasheet](http://www.atmel.com/Images/doc8161.pdf)). It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started. In order to fetch the data regarding the measurement of heart rate and Spo2,  a sensor named Max30100 is interfaced with Arduino uno.

As the Max30100 sensor does not work on a 3.3v or a 5v power supply, we need to add pull up resistors of the value 4.7 ohms along the breadboard to  ensure the efficient woking of the sensor module.

To add the instructions regarding how to sensors are supposed to work in this any environment, Arduino IDE is used. The open-source Arduino Software (IDE) makes it easy to write code and upload it to the board. It runs on Windows, Mac OS X, and Linux. The environment is written in Java and based on Processing and other open-source software. This software can be used with any Arduino board.Oled display screen is used to depict the readings fetched by the sensor.

After the data is fetched from the sensors, it is imparted to the application where other factors related to the heart diseases are provided  by the user.

**3.5.2 Non-Functional Requirements:**

* **Performance**

Performance can be directly stated as the response time. A better performance of the system is ensured if the response time to predict the answer after the data is fed by the user is as minimum as possible.

* **Scalability**

Scalability ensures that the system can effectively and efficiently predict the results regarding a heart disease detecting for N number of customers.

* **Usability**

Usability property ensures that the user interface of the system can easily be navigated by the user and overall usage of the model remains less tedious.

* **Availability**

Data can be accessed at all times since system is online. Background processes run as per designed and don’t affect user access.

* **Reliability**

Reliability property must ensure that the system provides accurate and reliable measurements for heart rate and Spo2 which are calculated by using a Arduino module and along with the individual values it must also ensure that the system provides accurate predictions depending on the data entered.

* **Recoverability**

The data about any user can be recovered from the database which is used to store the registrations details of the user which includes the username and password of the registered user.

* **Maintainability**

Any changes in the developed system can be maintained and managed without affecting any other component of the developed system.

* **Security**

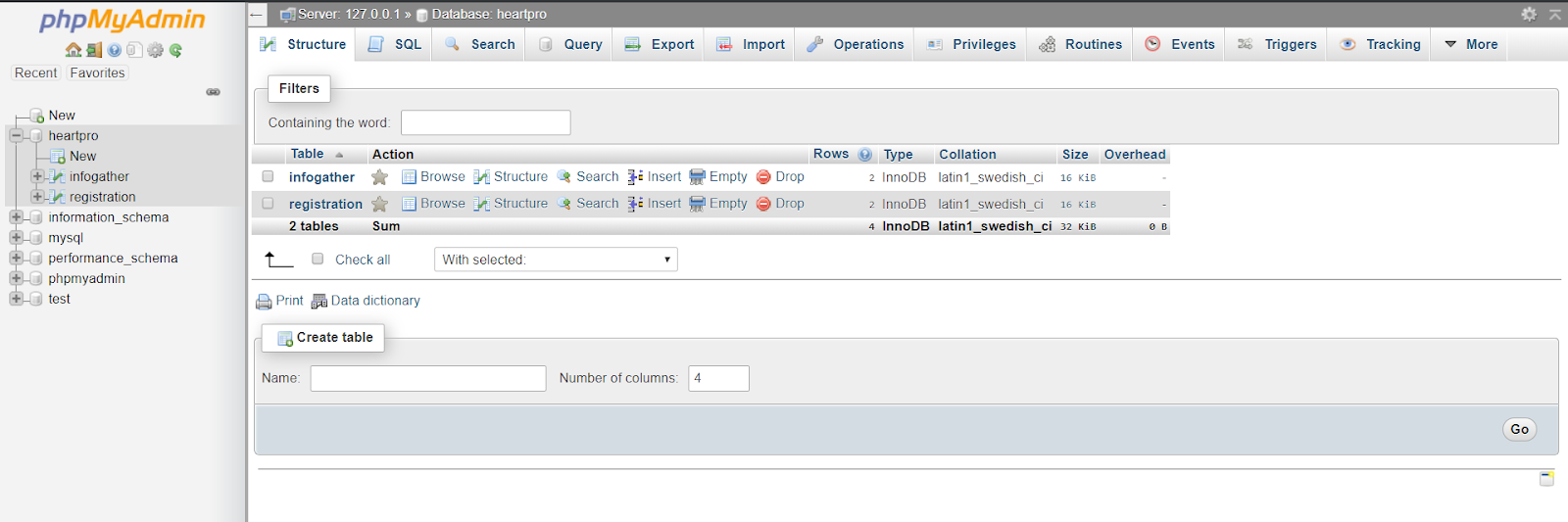
Security measures added in the system ensures that only a registered user can access the system to check his cardiac health. In order to register, user needs to provide his username along with a password which matches the prescribed password requirements.

**3.5.3 Interface Requirements**

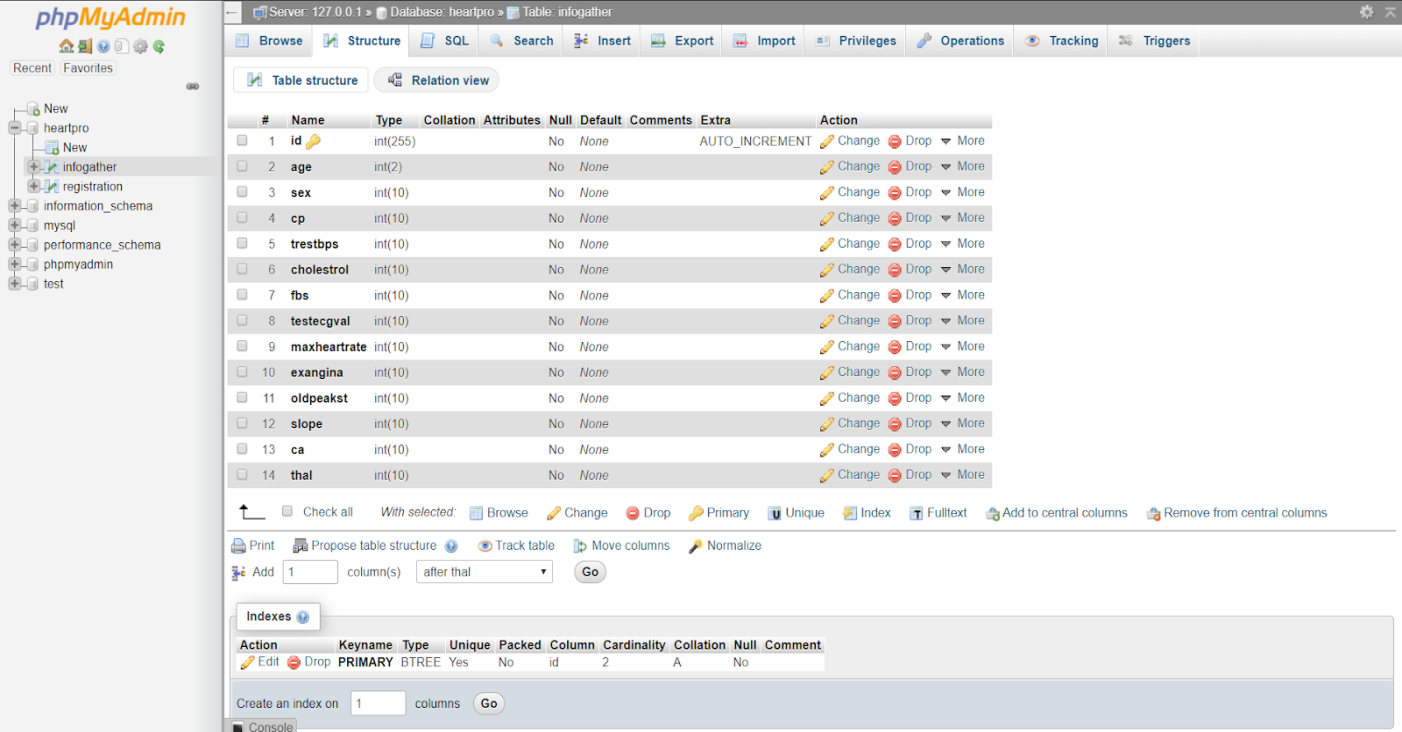
    The interface between the Arduino Uno and Max30100 is done by using a breadboard to mount the sensor module on it. As Max30100 sensor does not  support 3.3v or 5v power supply, various 4.7 ohms resistors are added to the sensor module for accurate functioning at available power supply. The display unit is being interfaced on the Arduino uno by using an SSD1306 OLED display. The data entered into the mobile application will be stored into a MySQL database as soon as the user submits the form. Using Ajax on the front end of the application the data is passed into a php framework which then enters the data into the database. On the back end the python file containing the algorithm fetches the data from the database and calculates the output. This is done with the help of Flask API which then the communicates with the html file and displays the output.

**3.6 Database Design**

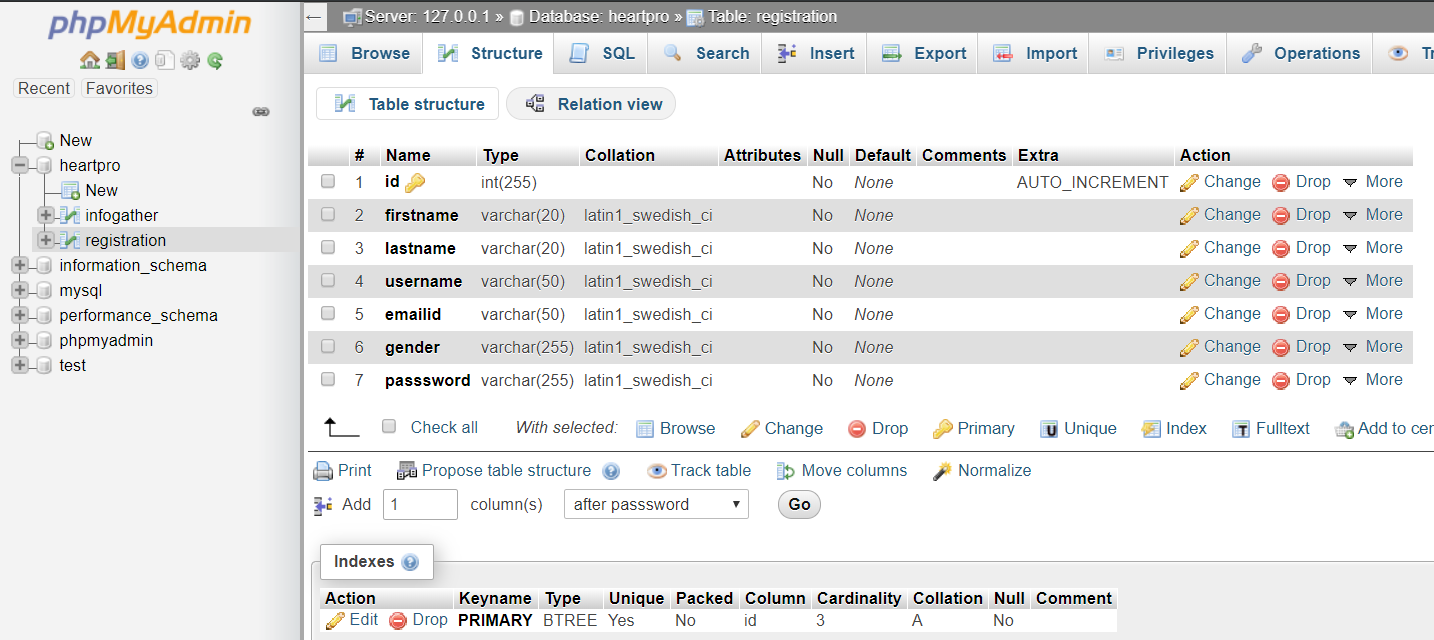
    The data is entered into a MySQL database once the user enters all the details. If any of the required fields are empty then the data of any other valued field will not be entered into the database. Every value plays a crucial role. The details entered for registration and form the medical report are kept in two different tables.  This is important to do because the machine learning algorithm takes the medical values as an input.



**Fig. 3.6.1: Database Overview**

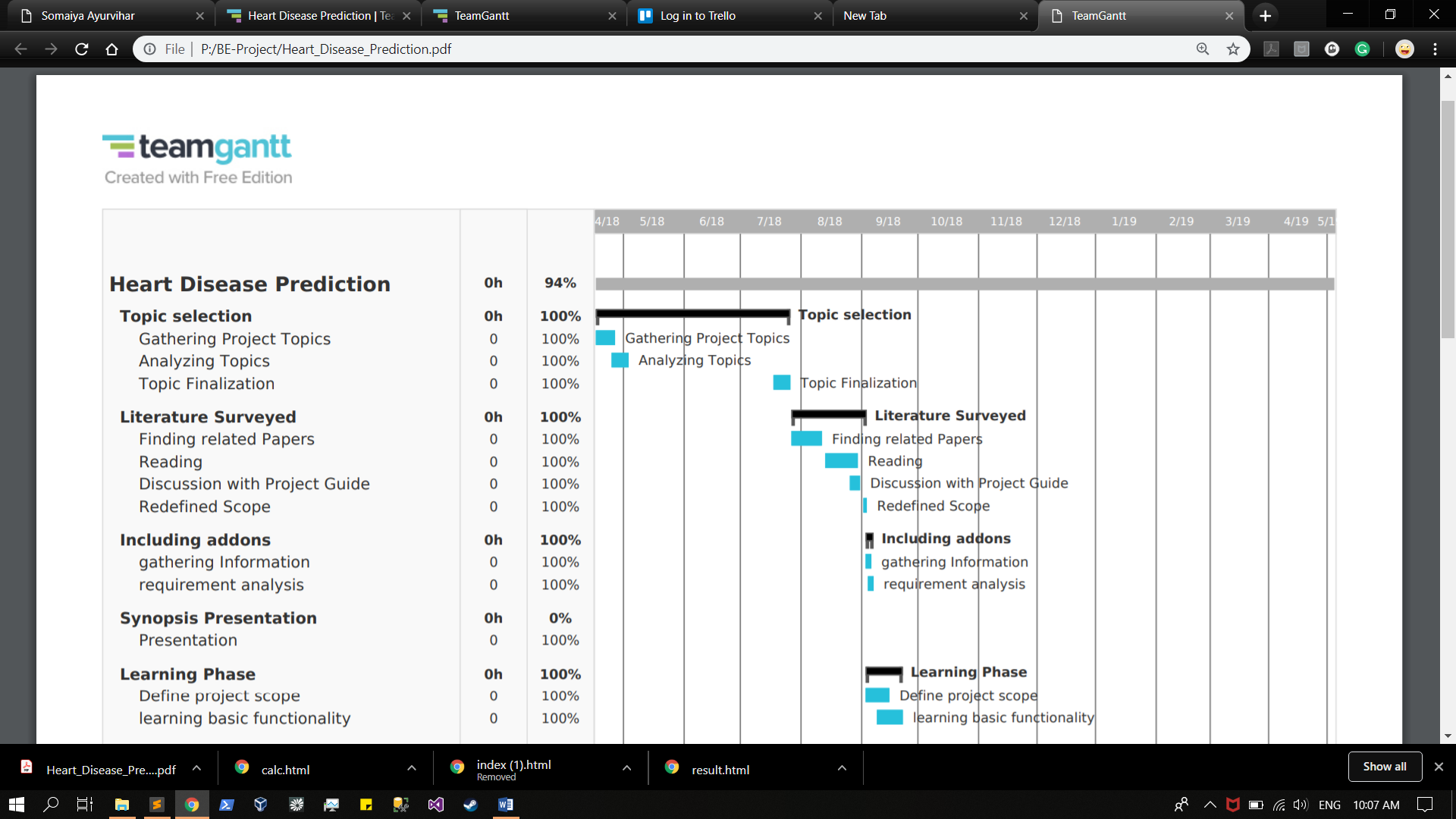


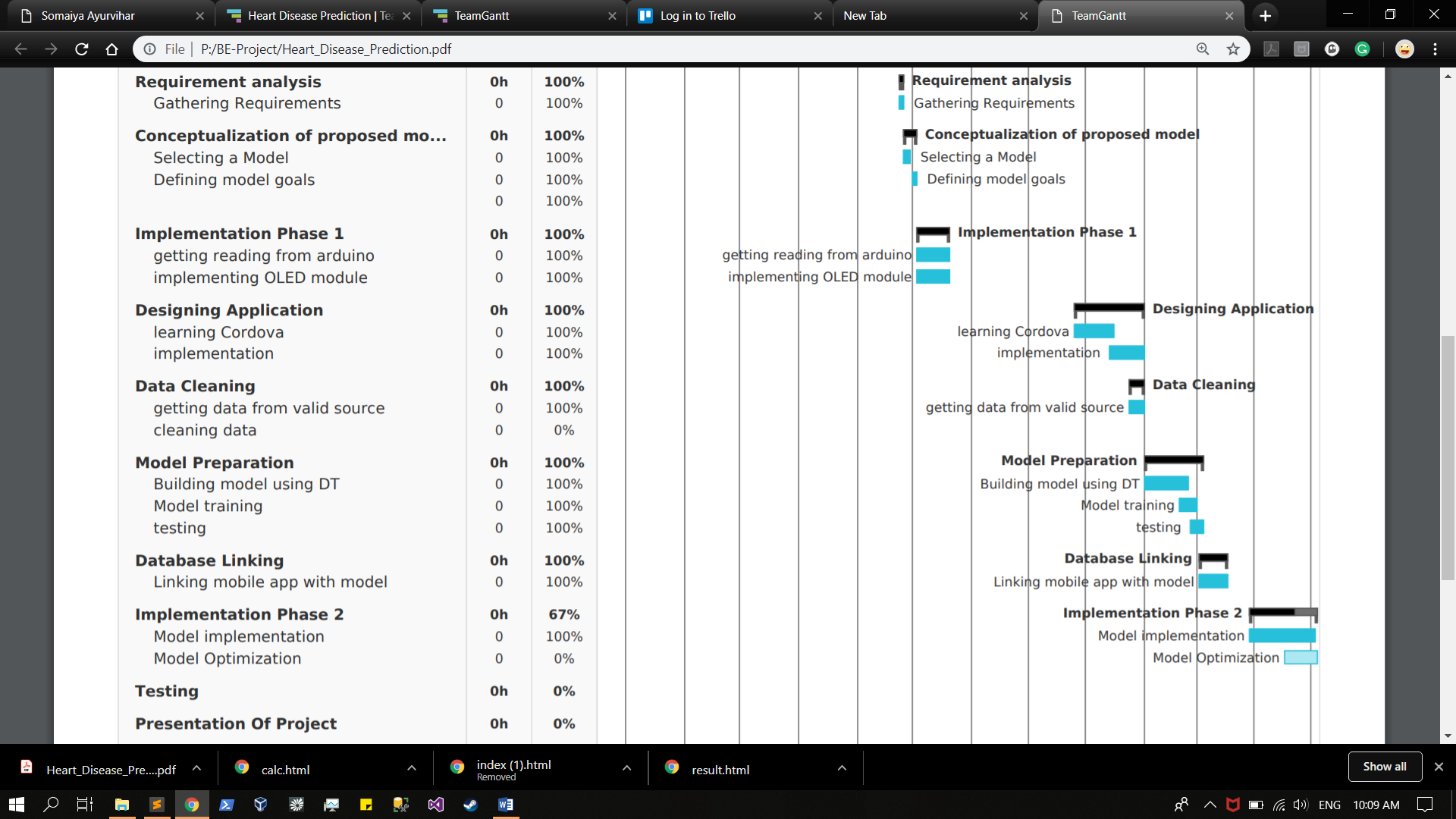
**Fig. 3.6.2: Table structure of the medical values**



**Fig. 3.6.3: Table structure of registration details**

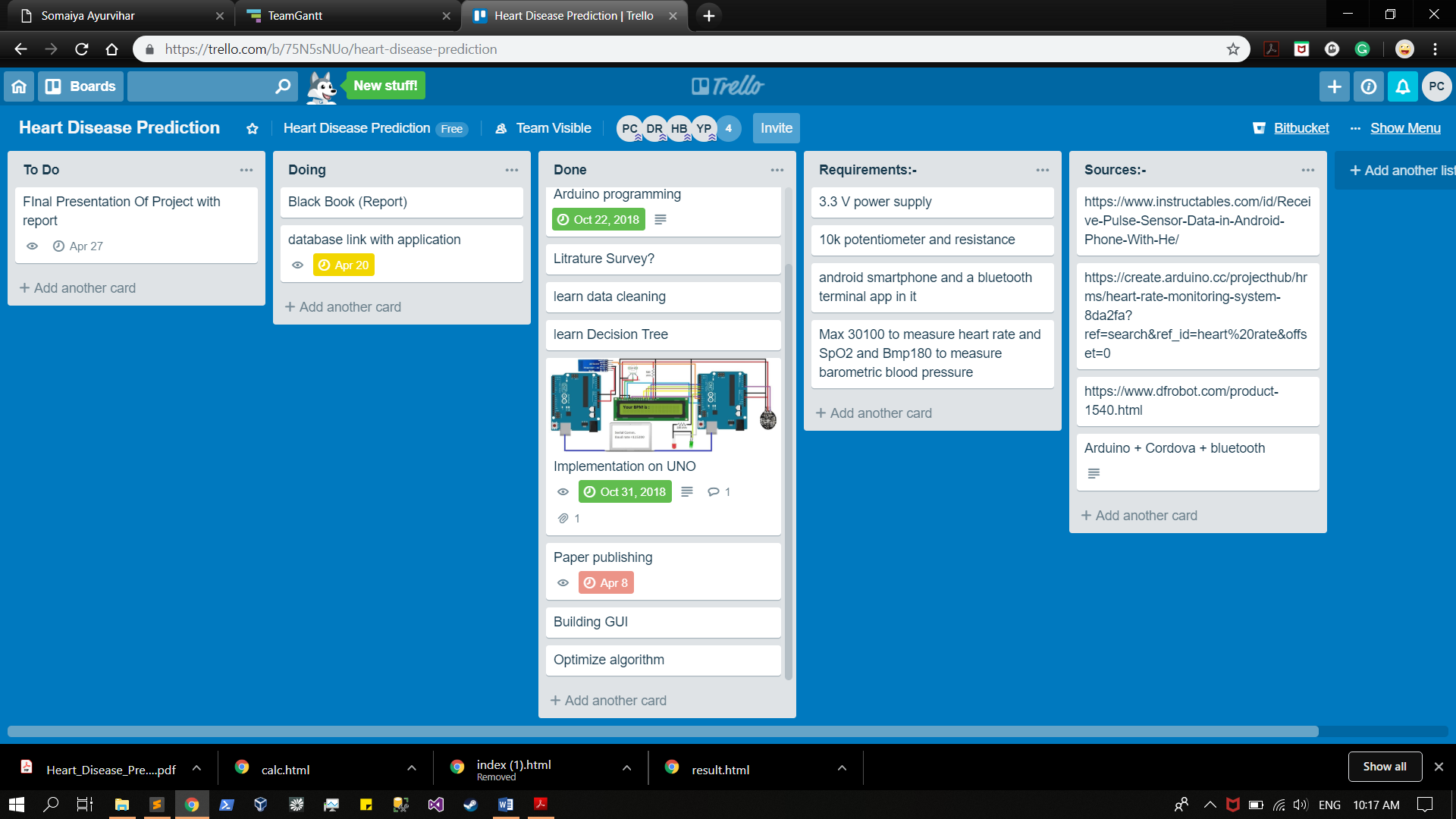
3.7 Project Scheduling





**Fig.3.7.1: Gantt chart representation**

Fig.3.7.1, depicts the project scheduling which is practiced by using teamgantt software.



**Fig.3.7.2: Project management**

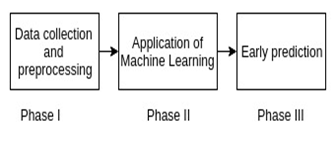
**Fig.3.7.2, depicts the screenshot of the project management software Trello, using is used to segregate project progress into three phase which are To do phase, Doing phase and Done phase.**

**CHAPTER 4**

**PROPOSED WORK**

**4.1 Proposed Approach**

The proposed application is as follows:



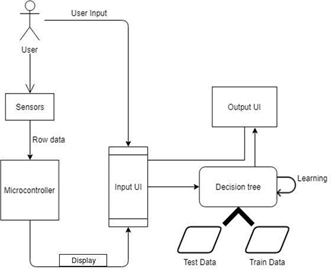
**Fig. 4.1.0: Modular view of proposed system**

In Phase I, the proposed system performs the task of data collection and preprocessing. Phase I exhibits actual implementation of Arduino UNO along with MAX30100 sensor which is used to obtain real time data which comprises of the parameters like heart rate and SpO2. In order to obtain the data, user needs to place the index finger over the MAX30100 module which then captures the data with the help two LEDs, one emitting a red light, another emitting infrared light. The data captured by the sensor can be viewed using a serial monitor present in the Arduino IDE. Collected data is then passed into Phase II to train the machine in order to obtain predicted results.

In Phase II, The proposed system uses python, which is widely used for machine learning as it has simple syntax unlike C++ and java. Libraries like Scikit  learn, Pandas, Numpy gives a great hand to build a machine learning algorithms. Jupyter notebook on the other end gives interactive cell structure for data cleaning, data visualization to machine learning. Data preprocessing is done on the Jupyter notebook also the functions like cleaning the data, finding redundant entries, removing null values can be done easily with the help of Pandas and Jupyter notebook. This preprocessed data is now ready to train a machine, to give the best accuracy on given dataset. The advantage of using the decision tree is that it gives out the feature that have caused the disease to occur. Information gain on all of the features is calculated and the highest gain of information gives the first split of the decision tree, based on that feature. To reduce the effect of bias resulting from the use of information gain, a variant known as the gain ratio is used. The information gain may have taste toward several outcomes. That is, it prefers to select attributes having a large number of values. Gain magnitude relation adjusts the data

gain for every attribute to permit for the breadth and uniformity of the attribute values. The algorithm gives an accuracy of 71% on 270 entries of data with 14 features. The input data which come from the user then get passed to the machine which gives the predicted results and also the main advantage of using the decision tree is that it gives out the feature that have caused the disease to occur.

In Phase III, with the help of a mobile application the data entered by the user as well as the real time data captured from Phase I is then passed to a trained machine in the form of a data frame. The user enters the data based on the data gained from the display and from the medical reports. Using the mobile application the user enters the data  into the algorithm. The algorithm is present in the back-end of the application. The Python libraries are used to integrate the machine learning algorithm with the mobile application. Entering data to the application is effortless. The user interface is simple to understand and hence it makes the application easy to use for general public. Depending on the data, machine predicts the result which is displayed on the output user interface. Further users are notified for an early diagnosis of heart disease by using alerts.



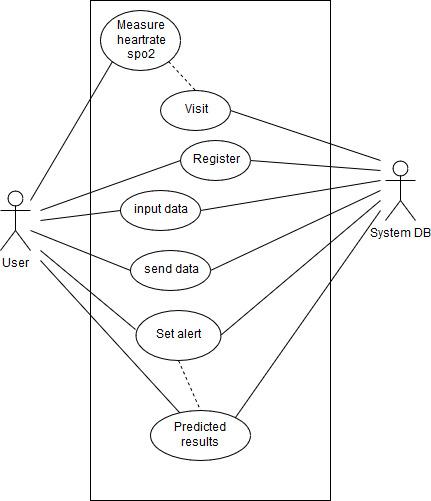
**Fig.4.1.1: Proposed approach**

Fig.4.1.1, depicts the system block diagram which explains the data flow through the system from its initial phase to the final phase of the process.

As per the Fig.4.1.0, in order to get the predicted results, the user initially needs to measure the reading of his heart rate and Spo2 by placing his index finger on the sensor module interfaced on the Arduino. The values captured by the sensor are then passed on to the microcontroller where these values get serialized. These serialized values are displayed by using OLED display.

The user is then directed to a user interface where he needs to enter other parameters which are used by the decision tree algorithm to give predictions which states whether the user will suffer from a cardiac disease or not. The decision tree uses two data sets, namely, test dataset and train dataset. The train dataset is used to train the module to carry out efficient predictions. The predicted results are displayed on the output interface.

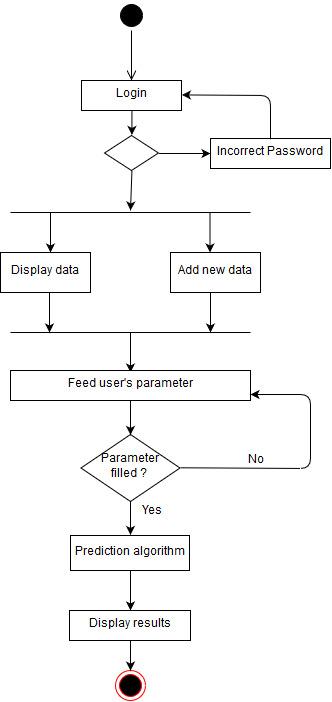
**4.2 Use case diagram**



**Fig 4.2: Use case diagram**

Fig.4.2.0, depicts the use case diagram which comprise of a set of actions, services, and functions that the system needs to perform.

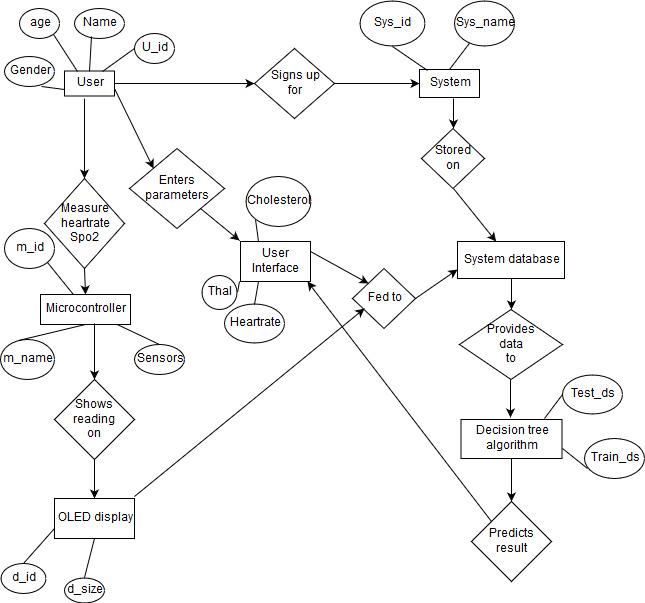
**4.3 Activity Diagram**



**Fig.4.3: Activity diagram**

Fig.4.3, depicts the activity diagram which represent the flow from one activity to another activity. The activity can be described as an operation of the system.

**4.4 Entity Relationship diagram**



**Fig.4.4: Entity Relationship diagram**

Fig.4.4, depicts the Entity relationship diagram which is based on the notion of real-world entities and the relationship between them.

**4.5  Product Cost and Expenses**

| **Component Name** | **Price** |
| --- | --- |
| Arduino UNO | 510\-₹ |
| Max30100 sensor | 500\-₹ |
| Breadboard | 30\-₹ |
| Connecting wires | 50\-₹ |
| 4.7 ohms resistors | 20\-₹ |
| OLED display | 350\-₹ |
| Power supply cable | 100\-₹ |

**CHAPTER 5**

**IMPLEMENTATION DETAILS AND RESULTS**

**5.1 Implementation details:**

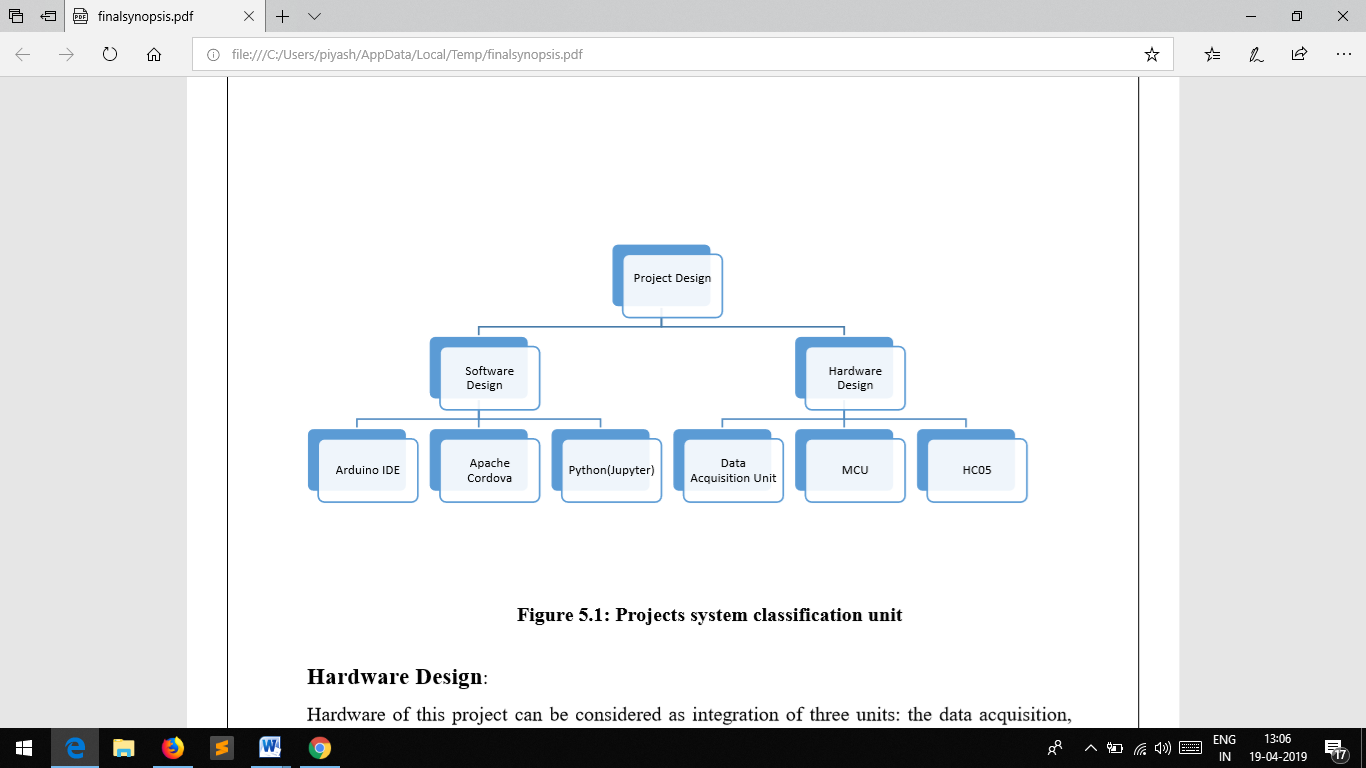
**Fig.5.1.0: Implementation details**

Fig.5.1.0, depicts the hardware implementations which include the hardware requirements as well as software requirements for the fulfilment of the project

**Hardware Details:-**

* Arduino Uno ATMEGA328p
* MAX30100 SENSOR
* Breadboard
* Connecting wires
* 4.7 ohm resistors
* OLED display

**Software Details:-**

* Arduino ide
* Python and required libraries
* Jupyter notebook
* Apache Cordova

**5.1.1 Hardware Design**:

Hardware of this project can be considered as integration of three units: the data acquisition, the MCU, display unit.

1. **Data Acquisition Unit:**

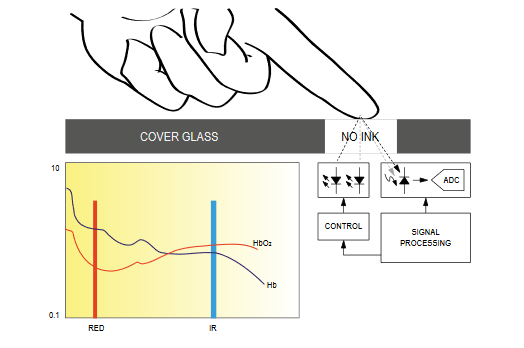
This unit is mainly responsible for obtaining patient’s vital parameters utilizing sensors. Sensors are devices that detect the variations and mainly are two types of them, Optical and solid state sensors.The sensors used in this project is a Max30100 sensor which is used to measure patient’s heart rate as well as the Spo2 level.

**Max30100 sensor:** The MAX30100 is an integrated pulse oximetry and heart-rate monitor sensor solution. It combines two LEDs, a photodetector, optimized optics, and low-noise analog signal processing to detect pulse oximetry and heart-rate signals. The MAX30100 operates from 1.8V and 3.3V power supplies and can be powered down through software with negligible standby current, permitting the power supply to remain connected at all times.



**Fig.5.1.1: Max30100 sensor**

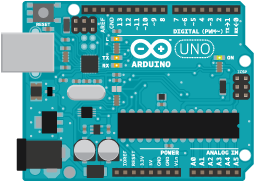
Fig.5.1, depicts the Max30100 sensor used to measure heart rate as well as Spo2 level in user’s body



**Fig.5.1.2: System Block Diagram**

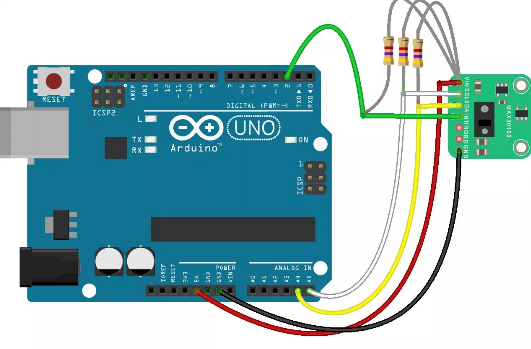
Fig.5.1.2, depicts the system block diagram of Max30100 sensor which shows the entire working of Max30100 sensor.

1. **Microcontroller (MCU):** The filtration of the measured signals, apply calculation on it and prepare it for transmission to the next unit mainly done by the microcontroller unit. In this project, the microcontroller chosen is a UNO Arduino board which is based on ATMEGA328 controller. Arduino Uno is a small and complete board with the same functionality of Arduino Duemilanove just in different package. The Arduino UNO is an open-source microcontroller board based on the [Microchip](https://en.wikipedia.org/wiki/Microchip_Technology) [ATmega328P](https://en.wikipedia.org/wiki/ATmega328P) microcontroller and developed by [Arduino.cc](https://en.wikipedia.org/wiki/Arduino). The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. The board has 14 Digital pins, 6 Analog pins, and programmable with the [Arduino IDE](https://en.wikipedia.org/wiki/Arduino#Software) (Integrated Development Environment) via a type B USB cable. It can be powered by a USB cable or by an external 9 volt battery, though it accepts voltages between 7 and 20 volts. It fits perfectly for this project as it’s easy to use as well as it provides mobility feature due to its suitable size for a wearable device.



**Fig.5.1.3: Arduino uno**

Fig.5.1.3, depicts the diagram of an Arduino microcontroller with all its pins specified appropriately.



**Fig. 5.1.4: Arduino interface with MAX30100**

In Fig.5.1.4 Arduino UNO is shown interfaced with MAX30100 sensors with the help of three 4.7 ohms resistors which acts as a pull up resistors to provide adequate power supply to the sensor.

**C) Display unit:** The term OLED stands for “*Organic Light emitting diode”* it uses the same technology that is used in most of our televisions but has fewer pixels compared to them. Here we are using SSD1306 OLED display which is interfaced to arduino in order to present the measurements of the heart rate and Spo2 on the display screen.



**Fig.5.1.5: SSD1306 OLED display**

Fig.5.1.5, depicts the OLED display used in the project to interfacr with arduino to present the reading of max30100 sensor on the display screen.

**Working of an OLED display:** In order to make something appear on the OLED screen we communicate with the **SSD1306 IC** present in the OLED module. This SSD1306IC will then update each pixel present on our OLED display. This communication can happen via IIC or SPI from any Microcontrollers like Arduino, PIC, etc. In order to communicate with an IC through any of the communication protocol we should first understand the IC by reading its datasheet which is a tiring but useful method. Here we have interfaced [OLED with Arduino](https://circuitdigest.com/microcontroller-projects/arduino-ssd1306-oled-display). There are lot of Libraries available for interfacing it with different Microcontrollers, using which we can make the interfacing a lot simpler. These libraries are easy to use and have lot of readily available graphical options. Also there are many online tools available for converting an image into a bitmap values to be fed into microcontrollers.

**5.2 Software Design:**

Several software tools were used throughout the entire development procedures of this project in order to program the Arduino board which is considered the core of this project besides developing the android application which will detect and alert when suspecting a heart attack.

**5.2.1 Arduino IDE:** Is the required software environment to program the Arduino by writing a code and upload it to the Arduino Uno. It also outputs the results for analysis using both serial monitor and serial plotter. The version used in this project is 1.8.3 which supports both serial monitor to print the HR wave while the serial monitor to print the temperature values. The Arduino IDE used to write a code to the Nano Arduino that has three main functions. These functions are: measuring BPM, measuring body temperature, sending measured data and alert when detect an abnormality.

**5.2.2 Python Libraries:** Scikit-Learn is a machine learning library for python and is designed to interoperate with the scientific and numerical libraries of python such as SciPy and NumPy. It is majorly considered for bringing machine learning into a production system.Scikit-learn offers a range of unsupervised and supervised learning algorithms through a consistent interface in python. Scikit-learn is built upon the SciPy. So, before you can use scikit-learn, you have to install SciPy. Modules or extensions for SciPy are conventionally named as SciKits. The module, as such, provides learning algorithms and is known as scikit-learn. TensorFlow is a famous open source deep learning library developed by the Google Brain team within the Machine Intelligence Research organization of Google. It is a blend between network specification libraries such as Lasagne and Blocks, and symbolic computation libraries such as Theano. If you are using Google voice search or Google photos, then you are utilizing the models built with Tensorflow. Tensorflow is a computational framework used for the purpose of expressing algorithms that involve numerous Tensor operations. As neural networks are expressed in the form of computational graphs, their implementation is done using Tensorflow in the form of a series of operations on Tensors. Tensors are nothing but the N-dimensional matrices that represent our data.

**5.2.3 The Jupyter Notebook:** It is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more. The Notebook has support for over 40 programming languages, including Python, R, Julia, and Scala.

**5.2.4 Apache Cordova:** Apache Cordova (formerly PhoneGap) is a [mobile application development framework](https://en.wikipedia.org/wiki/Multiple_phone_web-based_application_framework) originally created by [Nitobi](https://en.wikipedia.org/wiki/Nitobi). Apache Cordova enables [software](https://en.wikipedia.org/wiki/Computer_software) [programmers](https://en.wikipedia.org/wiki/Programmer) to build [applications](https://en.wikipedia.org/wiki/Application_software) for mobile devices using [CSS3](https://en.wikipedia.org/wiki/CSS3), [HTML5](https://en.wikipedia.org/wiki/HTML5), and [JavaScript](https://en.wikipedia.org/wiki/JavaScript) instead of relying on platform-specific [APIs](https://en.wikipedia.org/wiki/API) like those in [Android](https://en.wikipedia.org/wiki/Android_(operating_system)), [iOS](https://en.wikipedia.org/wiki/IOS), or [Windows Phone](https://en.wikipedia.org/wiki/Windows_Phone). It enables wrapping up of CSS, HTML, and JavaScript code depending upon the platform of the device. It extends the features of HTML and JavaScript to work with the device. The resulting applications are hybrid, meaning that they are neither truly native mobile application (because all layout rendering is done via Web views instead of the platform's native UI framework) nor purely Web-based (because they are not just Web apps, but are packaged as apps for distribution and have access to native device APIs). Mixing native and hybrid code snippets has been possible since version 1.9. The software was previously called just "PhoneGap", then "Apache Callback". As [open-source software](https://en.wikipedia.org/wiki/Open-source_software), Apache Cordova allows wrappers around it, such as [Appery.io](https://en.wikipedia.org/w/index.php?title=Appery.io&action=edit&redlink=1) or [Intel XDK](https://en.wikipedia.org/wiki/Intel_XDK). PhoneGap is Adobe's commercial version of Cordova along with its associated ecosystem. Many other tools and frameworks are also built on top of Cordova, including [Ionic](https://en.wikipedia.org/wiki/Ionic_Framework), [Monaca](https://en.wikipedia.org/wiki/Monaca_(software)), [TACO](https://en.wikipedia.org/w/index.php?title=TACO&action=edit&redlink=1), Onsen UI, Visual Studio, GapDebug, App Builder, Cocoon, Framework7, Quasar Framework, Evothings Studio, NSB/AppStudio, Mobiscroll, the [Intel XDK](https://en.wikipedia.org/wiki/Intel_XDK), and the [Telerik Platform](https://en.wikipedia.org/wiki/Telerik). These tools use Cordova, and not PhoneGap for their core tools.

**5.2.5 Flask:** Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications. Flask is considered more [Pythonic](http://blog.startifact.com/posts/older/what-is-pythonic.html) than the [Django](https://www.fullstackpython.com/django.html) web framework because in common situations the equivalent Flask web application is more explicit. Flask is also easy to get started with as a beginner because there is little boilerplate code for getting a simple app up and running.

**5.2.6 Decision Tree:**

**Advantage 1: Decision trees implicitly perform variable screening or feature selection**

When we fit a decision tree to a training dataset, the top few nodes on which the tree is split are essentially the most important variables within the dataset and feature selection is completed automatically!

**Advantage 2: Decision trees require relatively little effort from users for data preparation**

To overcome scale differences between parameters - for example if we have a dataset which measures revenue in millions and loan age in years, say; this will require some form of normalization or scaling before we can fit a [regression model and interpret the coefficients](http://www.simafore.com/blog/bid/56752/3-checks-to-prevent-abuse-of-regression-models).  Such variable transformations are not required with decision trees because the tree structure will remain the same with or without the transformation. Another feature which saves data prep time: missing values will not prevent splitting the data for building trees. Decision trees are also not sensitive to outliers since the splitting happens based on proportion of samples within the split ranges and not on absolute values.

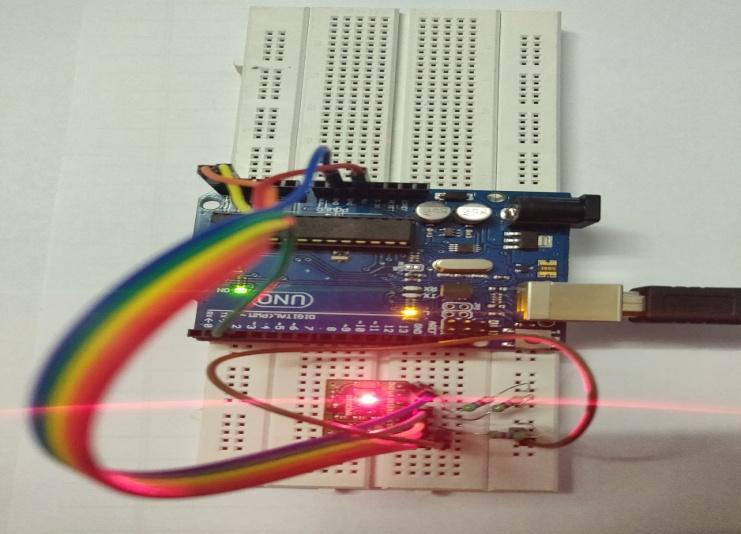
**Advantage 3: Nonlinear relationships between parameters do not affect tree performance**

As we described [here, highly nonlinear relationships](http://www.simafore.com/blog/bid/54963/6-checkpoints-to-ensure-regression-model-validity-for-analytics) between variables will result in failing checks for simple regression models and thus make such models invalid. However, decision trees do not require any assumptions of linearity in the data. Thus, we can use them in scenarios where we *know* the parameters are nonlinearly related.

**Advantage 4: The best feature of using trees for analytics - easy to interpret and explain to executives!**

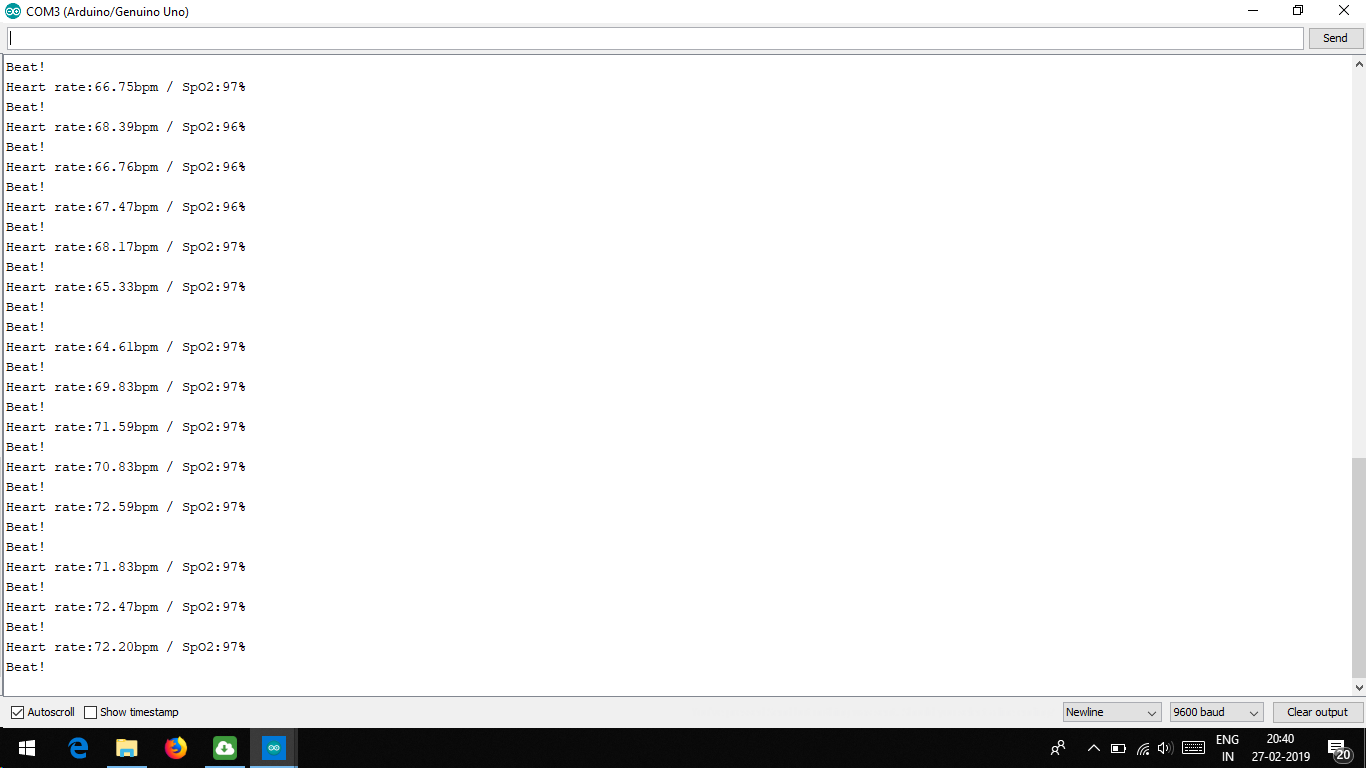
These advantages need to be tempered with one key disadvantage of decision trees: without proper pruning or limiting tree growth, they tend to overfit the training data, making them somewhat poor predictors.

**5.2 Results and Discussions**



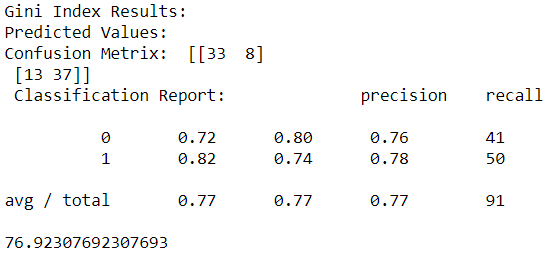
**Fig. 5.2.1: Actual IOT implementation**

Fig.5.2.1 depicts the implementation of Arduino along with Max30100 sensor which forms IOT module that is used to accumulate user data. The arduino module is being interfaced with a MAX30100 sensor by using a breadboard, if you look closely, the SCL and SDA pins are pulled-up via the 4.7k ohm resistors to 1.8V, due to which it won't work well with microcontrollers with higher logic levels. The solution to this problem is to remove the resistors from the board (encircled on the image below) and attach external 4.7k ohms resistors instead.



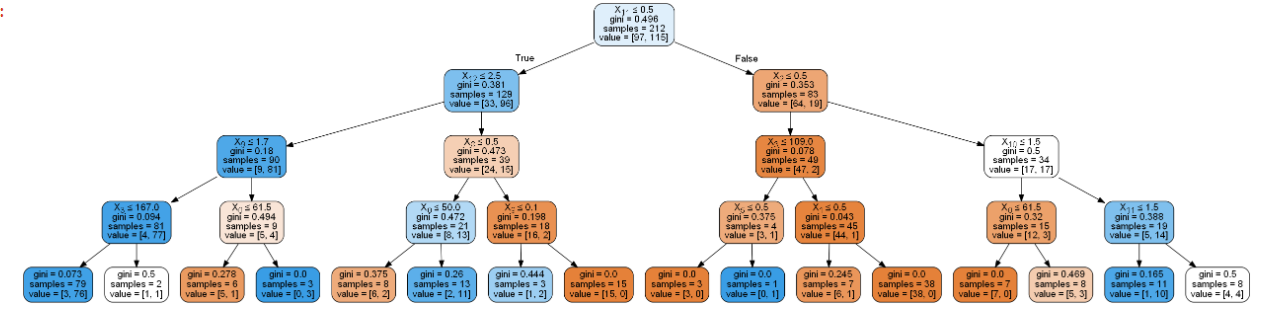
**Fig. 5.2.2: Output obtained from serial monitor**

Fig.5.2.2 depicts the output obtained from the Arduino microcontroller which comprises of parameters such as heart rate and SPO2.



**Fig. 5.2.3: Decision tree accuracy**

Fig.5.2.3 shows the results of algorithm accuracy with confusion matrix. From the above figure it is seen that the decision tree algorithm provides us with an accuracy of 76.92 percent.



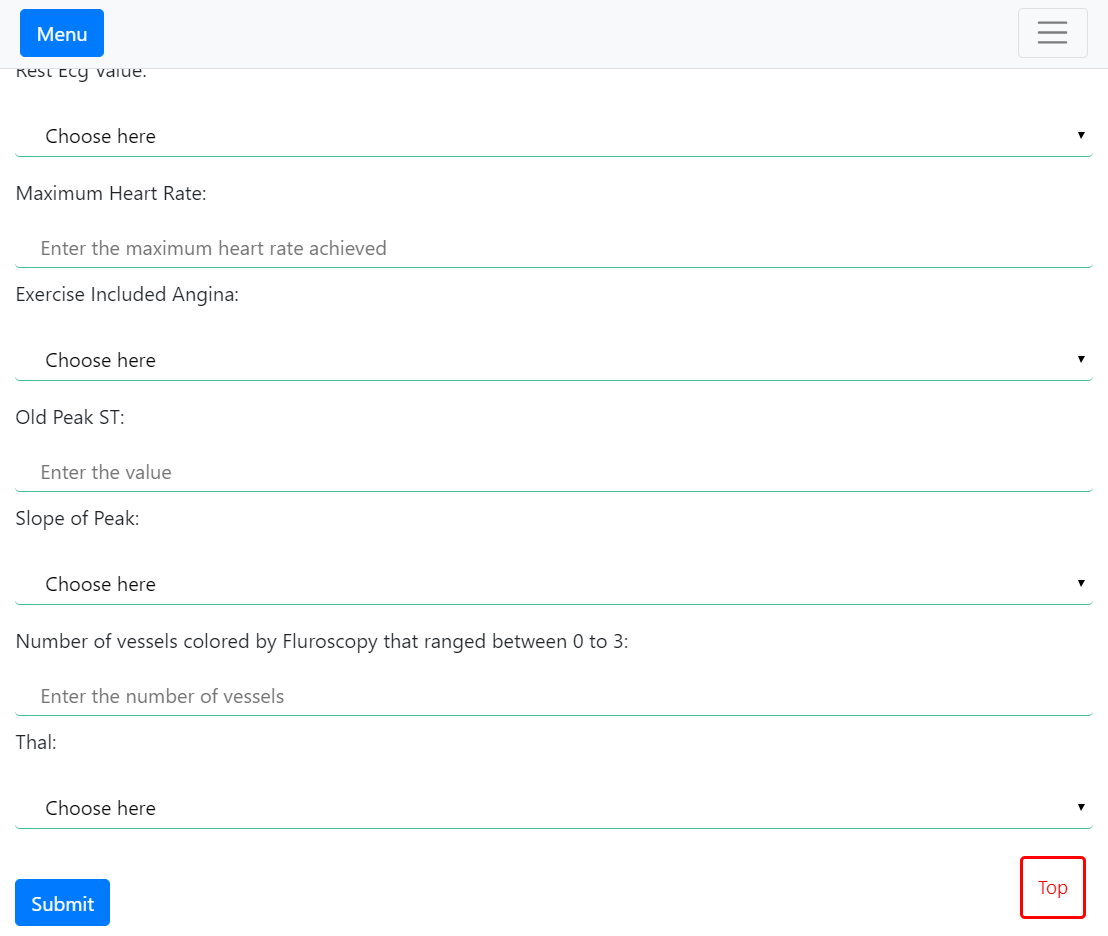
**Fig. 5.2.5: Decision tree**

Fig.5.2.5, depicts the decision tree indicating various splits followed in a tree from the root node to its leaf nodes



**Fig. 5.2.5: Oxywatch measurements**

Fig.5.2.5 depicts the actual output of Oxywatch device used by the doctors to calculate patients heart rate and Spo2.



**Fig. 5.2.6: User interface**

Fig.5.2.6, depicts the screen shot of the user interface which allows the user to enter the details which in turn are used to predict the result which stated whether or not the person will suffer from a heart disease.

**CHAPTER 6**

**CONCLUSION AND FUTURE WORK**

CVD is a major public health problem in India, often impacting the most productive years of an individual’s life. The epidemiological transition plays out differently in different regions of India because of varied economic development. Heart diseases when aggravated spiral way beyond control. Diseases related to heart are much more complex than other diseases and a large amount of people have lost their lives due to such diseases. In small period of time the patients have to face serious consequences if the early symptoms of heart diseases are ignored. Sedentary lifestyle and excessive stress in today’s world have worsened the situation. Thus, the early detection of such diseases assist in keeping them in under control. However, daily exercise and getting rid of unhealthy habits is advised. The odds of getting stroke and heart diseases grows due to consumption of tobacco and unhealthy diets. Hence, the work proposes a mobile application that uses human body parameters like heart rate and Spo2 retrieved from sensors as well as entered by user and applies decision tree classification algorithm in machine learning to predict risk of a cardiac disease.

The work can be extended to use real data of patients from hospitals across the nation. The data will have to be pre-processed to make it suitable for applying machine learning. Further, an intelligent system may be developed as a future work that can lead to selection of proper treatment methods for a patient diagnosed with heart disease. Heart beat classification is considered as the main tool for recognizing and diagnosing different heart diseases. The automation of heart beat classification is very necessary due to the exhaustive process of the 24 h mentoring of Electrocardiogram (ECG) signal of the heart. Moreover, ECG is considered as one of the most powerful tools for the diagnosis of heartbeats. Developing a model which trains the machine using the parameter of Electrocardiogram would also increase the accuracy of prediction of an early heart disease.

**CHAPTER 7**

**REFERENCES**

[1] S. Reddy K, B. Shah, C. Varghese, A. Ramadoss, “Responding to the threat of chronic diseases in India”, The Lancet, 2005. [2] V. Fuster, “Cardiac statistics”, Journal of the American College of Cardiology, Elsevier, 2016. [3] S. Spencer, R. Horton, “The Lancet Cardiology Collection”, The Lancet, 2012. [4] D. Prabhakaran, S. Yusuf, S. Mehta, J. Pogue, A. Avezum, A. Budaj, L. Cerumzynski, M. Flather, K. Fox, D. Hunt, L. Lisheng, M. Keltai, A. Parkhomenko, P. Pais, S. Reddy, M. Ruda, T. Hiquing, Z. Jun, “Two-year outcomes in patients admitted with non-ST elevation acute coronary syndrome: results of the OASIS registry 1 and 2”, Indian Heart Journal, 2005. [5]J. Yang, J. Kim, U. Kang, Y. Lee, “Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFIS-LDA)”, Personal and Ubiquitous Computing, 2013. [6]M. Masud, M. Serhani, A. Navaz, “Resource-Aware Mobile-Based Health Monitoring”, IEEE Journal of Biomedical and Health Informatics, 2017. [7] M. Gandhi, S. N. Singh, “Predictions in heart disease using techniques of data mining”, International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), 2015. [8] H. Elsayed, L. Syed, “An Automatic Early Risk Classification of Hard Coronary Heart Diseases using Framingham Scoring Model”,The Second International Conference on Internet of things, Data and Cloud Computing, 2017. [9] P. Rajbhandary, B. Zhou, “Detecting Heart Abnormality using ECG with CART”, unpublished, 2019. [10] L. Chen, Q. Cao, S. Li, X. Ju, “Predicting Heart Attacks”, International Journal of Computer Applications, 2018. [11] C. Ordonez, “Comparing Association Rules and Decision Trees for Disease Prediction”, in proceedings of the international workshop on Healthcare information and knowledge management, pp.17-24, 2016. [11] S. Fuicu, A. Avramescu, “Real time E-health system for continuous care” in proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, pp.436-439,2018. [12]M. Shouman, T. Turner, R. Stocker, “Using Decision Tree for Diagnosing Heart Disease Patients”, in proceedings of the Ninth Australasian Data Mining Conference, 2011. [13] R. Bialy, M. Salama, “An ensemble model for heart disease data sets: a generalized model” in proceedings of the 10th International Conference on Informatics and Systems, 2016. [14] C. Ordonez, “Association rule Discovery with Train and Test approach for Heart Disease Prediction”, Journal of Intelligent Data Analysis, 2011. [15] T. Tavares, A. Oliveira, G. Cabral, S. Mattos, R. Grigorio, “Processing unbalanced data using weighted support vector machine for prediction of heart disease in children”, in [International Joint Conference on Neural Networks (IJCNN)](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6691896), 2013. [16] A. Mdhaffar, I. Rodriguez, K. Charfi, L. Abid, B. Freisleben, “Complex event processing for heart failure prediction”, [IEEE Transactions on NanoBioscience](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=7728), 2017 [17] F. Miao, Y. Cai, Y. Zhang, X. Fan, Y. Li, “Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest ”, in [IEEE Access](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6287639), 2018. [18] E. Loukis, M. Maragoudakis, “Heart murmurs identification using random forests in assistive environments”, 3rd International Conference on Pervasive Technologies Related to Assistive Environments, 2010. [19] N. Allahverdi, S. Torun, I. Saritas, “Design of a fuzzy expert system for determining of coronary heart disease risk”, in Proceedings of International Conference on Computer Systems and Technologies, CompSysTech, 2007. [20] N. Khateeb, M. Usman, S. Zulfikar, “Efficient heart disease prediction system using K-nearest neighbor classification technique” in proceedings of the International Conference on Big Data and Internet of Thing, pp.406-409, 2017. [21] J. Ghosh, M. Valtorta, “Building a bayesian Network model of Heart disease” in proceedings of the 38th annual on Southeast regional conference, pp.239-240, 2000. [22] H. Kahtan, K. Zamli, “Heart disease diagnosis system using fuzzy logic” in proceedings of the 2018 7th International Conference on Software and Computer Applications, pp.297-301, 2018. [23] Pulse Oximeter and Heart-Rate Sensor IC for Wearable Health, MAX30100, http://www.datasheets.maxim integrated.com.