CHAPTER 1

INTRODUCTION

The cardiovascular system consists of the heart, blood vessels, and the approximately 5 liters of blood that the blood vessels transport. Responsible for transporting oxygen, nutrients, hormones, and cellular waste products throughout the body, the cardiovascular system is powered by the body's hardest-working organ — the heart, which is only about the size of a closed fist. Even at rest, the average heart easily pumps over 5 liters of blood throughout the body every minute. The heart's job is to pump blood around the body. The heart is located in between the two lungs. It lies left of the middle of the chest.[1].

Fig.1 - Abstract view of normal Cardio-vascular system of human.

1.1 Structure of the Heart-

The heart is a muscle about the size of a fist, and is roughly cone-shaped. It is about 12cm long, 9cm across the broadest point and about 6cm thick. The pericardium is a fibrous

covering which wraps around the whole heart. It holds the heart in place but allows it to move as it beats. The wall of the heart itself is made up of a special type of muscle called cardiac muscle [1].

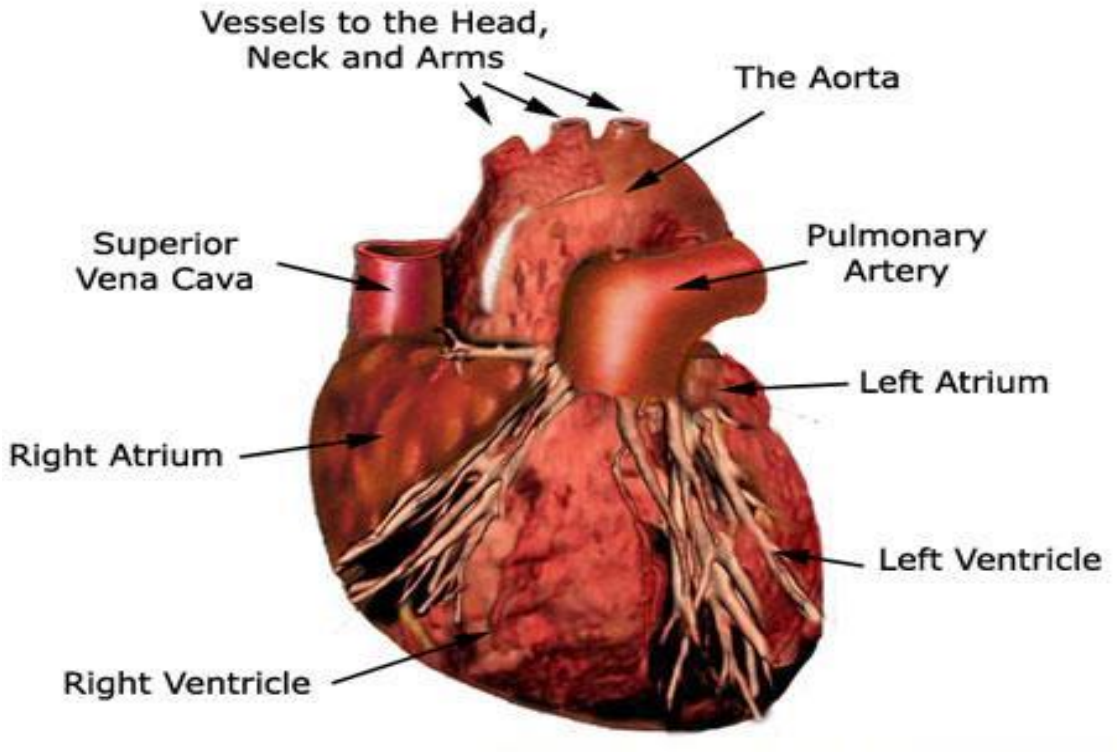


Fig.1.1 – The Heart

1.1.1 Chambers of the Heart-

The heart has two sides, the right side and the left side. The heart has four chambers. The left and right side each have two chambers, a top chamber and a bottom chamber. The two top chambers are known as the left and right atria (singular: atrium). The atria receive blood from different sources. The left atrium receives blood from the lungs and the right atrium receives blood from the rest of the body. The bottom two chambers are known as the left and right ventricles. The ventricles pump bloodout to different parts of the body.

The right ventricle pumps blood to the lungs while the left ventricle pumps out blood to the rest of the body. The ventricles have much thicker walls than the atria which allows them to perform more work by pumping out blood to the whole body [1].

1.1.2 Blood Vessels-

Blood Vessels are tubes which carry blood. Veins are blood vessels which carry blood from the body back to the heart. Arteries are blood vessels which carry blood from the heart to the body. There are also microscopic blood vessels which connect arteries and veins together called capillaries. There are a few main blood vessels which connect to different chambers of the heart. The aorta is the largest artery in our body. The left ventricle pumps blood into the aorta which then carries it to the rest of the body through smaller arteries. The pulmonary trunk is the large artery which the right ventricle pumps into. It splits into pulmonary arteries which take the blood to the lungs. The pulmonary veins take blood from the lungs to the left atrium. All the other veins in our body drain into the inferior vena cava (IVC) or the superior vena cava (SVC). These two large veins then take the blood from the rest of the body into the right atrium[1].

1.1.3Valves-

Valves are fibrous flaps of tissue found between the heart chambers and in the blood vessels. They are rather like gates which prevent blood from flowing in the wrong direction. They are found in a number of places. Valves between the atria and ventricles are known as the right and left atrioventricular valves, otherwise known as the tricuspid and mitral valves respectively. Valves between the ventricles and the great arteries are known as the semilunar valves. The aortic valve is found at the base of the aorta, while the pulmonary valve is found the base of the pulmonary trunk. There are also many valves found in veins throughout the body. However, there are no valves found in any of the other arteries besides the aorta and pulmonary trunk[1].

1.2The Cardiovascular System-

The cardiovascular system refers to the heart, blood vessels and the blood. Blood contains oxygen and other nutrients which your body needs to survive. The body takes these essential nutrients from the blood. At the same time, the body dumps waste products like carbon dioxide, back into the blood, so they can be removed. The main function of the cardiovascular system is therefore to maintain blood flow to all parts of the body, to allow it to survive. Veins deliver used blood from the body back to the heart. Blood in the veins is low in oxygen (as it has been taken out by the body) and high in carbon dioxide (as the body has unloaded it back into the blood). All the veins drain into the superior and inferior vena cava which then drain into the right atrium. The right atrium pumps blood into the right ventricle. Then the right ventricle pumps blood to the pulmonary trunk, through the pulmonary arteries and into the lungs. In the lungs the blood picks up oxygen that we breathe in and gets rid of carbon dioxide, which we breathe out. The blood is becomerich in oxygen which the body can use. From the lungs, blood drains into the left atrium and is then pumped into the left ventricle. The left ventricle then pumps this oxygen-rich blood out into the aorta which then distributes it to the rest of the body through other arteries[1].

Following functions are performed inside our cardiovascular system.

- Carotid arteries, which take blood to the neck and head
- Coronary arteries, which provide blood supply to the heart itself
- Hepatic artery, which takes blood to the liver with branches going to the stomach
- Mesenteric artery, which takes blood to the intestines
- Renal arteries, which takes blood to the kidneys
- Femoral arteries, which take blood to the legs
- The body is then able to use the oxygen in the blood to carry out its normal functions. This blood will again return back to the heart through the veins and the cycle continues.

1.3CARDIOVASCULAR DISEASES(CVD)-

Cardiovascular disease (CVD) is a class of diseases that involve the heart or blood vessels. Cardiovascular disease includes coronary artery diseases (CAD) such as angina and myocardial infarction (commonly known as a heart attack). Other CVDs include stroke, heart failure, hypertensive heart disease, rheumatic heart disease, cardiomyopathy, heart arrhythmia, congenital heart disease, valvar heart disease, carditis, aortic aneurysms, peripheral artery disease, thromboembolic disease, and venous thrombosis. [2]

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. [2]

1.3.1 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. [3]

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. [3]

1.3.2 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 buildup narrows the coronary arteries, and this causes the heart to receive less oxygen and nutrients. [3]

1.3.3 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.[4].

1.3.4 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[4].

1.3.5 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 [4].

1.3.6 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. [4].

1.3.7 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. People with this type of heart condition often feel tired and out of breath[4].

1.3.8 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.[4]

1.3.9 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. [5]

The underlying mechanisms vary depending on the disease. Coronary artery disease, stroke, and peripheral artery disease involve atherosclerosis. This may be caused by high blood pressure, smoking, diabetes, lack of exercise, obesity, high blood cholesterol, poor diet, and excessive alcohol consumption, among others. High blood pressure results in 13% of CVD deaths, while tobacco results in 9%, diabetes 6%, lack of exercise 6% and obesity 5%. Rheumatic heart disease may follow untreated strep throat. [5]

1.3.10 CVD Prevention-

It is estimated that 90% of CVD is preventable. Prevention of atherosclerosis involves improving risk factors through: healthy eating, exercise, avoidance of tobacco smoke and

limiting alcohol intake. Treating risk factors, such as high blood pressure, blood lipids and diabetes is also beneficial. Treating people who have strep throat with antibiotics can decrease the risk of rheumatic heart disease. The use of aspirin in people, who are otherwise healthy, is of unclear benefit. [5]

Current Scenario-

Cardiovascular diseases are the leading cause of death globally. This is true in all areas of the world except Africa. Together they resulted in 17.9 million deaths (32.1%) in 2015, up from 12.3 million (25.8%) in 1990. Deaths, at a given age, from CVD are more common and have been increasing in much of the developing world, while rates have declined in most of the developed world since the 1970s. Coronary artery disease and stroke account for 80% of CVD deaths in males and 75% of CVD deaths in females. Almost 19% of deaths in India in the year 2015 were attributed to Cardiovascular Diseases (CVD). Early detection and prediction of CVD is very important for patients' treatment and doctors' diagnose which can help to reduce mortality. Computational intelligence and Data mining plays an important role in the field of heart disease prediction. Thus, there arises a need to develop a support system for detecting heart diseases in a patient. Unexpected acute events have resulted in much affliction as well as high treatment costs. The latter are now reaching unsustainable levels and are becoming huge burdens even for developed countries. Early prediction and intervention would therefore be of huge benefit to society. In this paper, we propose efficient data mining techniques toselect the best features withthe lowest costs and shortest times and machine learning algorithms to achieve the accuracy. This technique will help to reduce the work load and cost for patients as well as health care unit. [6]

We are living in a postmodern era and there are tremendous changes happening to our daily routines which make an impact on our health positively and negatively.[7] As a result of these changes various kind of diseases are enormously increased. Especially, heart disease has become more common these days. The life of people is at a risk.[7]

Heart disease is the leading cause of death for both men and women. It is a serious disease since we often hear that most of the people die out of Heart diseases and other kinds of

similar diseases relates to heart.[8] It is observed by most of the medical scholars that at many times most of the heart patients might not survive heart attacks and they die with it. Know the warning signs and symptoms of a heart attack so that you can act fast if you or someone you know might be having a heart attack. The chances of survival are greater when emergency treatment begins quickly. The World Health Organization (WHO) analyzed that thirteen millions of death worldwide due to the reason of Heart diseases in 2017.[8]

We have an opportunity to provide early diagnoses and treatments for people who are likely to have heart failure and help them have longer, more active lives. A preferred strategy to resolve the problems of accurate diagnosis and the delivery of targeted therapies is the frequent performance of complete physical evaluations. [9] However, complete and frequent physical evaluations would lead to data overload. Heart failure patients and society would benefit if we could provide an accurate, systematic diagnostic service for the population. [9]

Heart failure, also referred to as congestive heart failure, occurs when the heart cannot pump enough blood to meet the body's needs. If the blood circulation to the body is insufficient, the organs of the body that is brain and heart stop working and death occurs in few minutes. [10]The risk parameters associated with heart failures are age, family history, hypertension, high cholesterol, diabetes, smoking, tobacco, alcohol consumption, poor diet and chest pain. Even about sixty percent of total population are suffering from the heart disease, so prediction of the heart disease earlier can prevent the heart disease.In many cases heart disease almost noticed at the final stage or after death. It’s difficult to cure heart disease at the final stages so people are very hesitant to treat at the early stages of heart disease.[10]

Heart failure is more common among people over the age of 65, overweight people, and those with a previous heart attack. The diagnostic method for heart failure is primarily based on the patient's medical and family histories, a physical examination, and test results. [11]The signs and symptoms of heart failure are also common in other conditions. Thus, physicians identify any damage to a patient's heart and check how well the patient's heart

pumps blood. [12]These diagnostic methods provide massive sequential data, and it is a non-trivial task to perform accurate diagnosis with such massive data, particularly in the early stages. Indeed, a method for early diagnosis of heart failure that has a low-error rate is critically needed for clinical trials and treatments.[13]

As per the Indian Heart Association, 50% of heart attack occur under 50 years of age and 25% of all heart attack occur under 40 years of age in Indians. Urban population is thrice as vulnerable to heart attacks as rural population. We thus propose to collect relevant data elements related to our field of study, train the data as per the proposed algorithm of machine learning and predict how strong is there a possibility for a patient will have a heart disease.[14]

In these days Data Mining is getting to be noticeable famous in healthcare field. Just because nowadays there is need of proficient investigative techniques in healthcare industries. Data mining is a procedure of examining information from alternate point of view and turnout the knowledge from it. Data mining is one of the tool which can be used for many of disease prediction such as heart disease, breast cancer and lung cancer etc. The objective of this review paper is to analyze the different Data mining techniques for predicting the heart disease prediction [14].

Many of hazards factors for heart disease for example, age, sexual orientations, hypertension, smoking, family history etc. Heart disease involves various types of conditions that can influence the heart work in which types including heart strokes, heart failure, heart attacks, heart valve disease, cardiovascular disease which is leading cause of death over past few years. Clinical choices are frequently made of in views of specialist and experience rather than to on the knowledgeable data information covered up in database. Patients and doctors need genuine detailed information about the chance of growing heart disease.[15]

DataMining joins measurable examination, machine learning and database innovation to separate hidden paradigm and connections from huge database. They utilize distinctive methodologies for producing the grouping models, which expands the odds for finding a prediction display with high characterization exactness. Decision tree algorithm recursively

separate perceptions in branches to develop a tree with the end goal of enhancing the expectation exactness. Neural Network is systematic method in which subjective framework and neurological elements of brain and capable predicting new observations from different perceptions in the wake of executing a procedure of existing information. [16]

Heart disease is a general name for a variety of diseases, whose symptoms may vary depending on the specific type of heart disease. The hospitals use database systems to store and manage their patient data. These systems generate large volumes of data, but these data are rarely used to support perceptive clinical decision making. Big data coupled with data mining algorithms makes it possible to do a large number of things such as, identify healthcare trends, disease prevention, and early diagnosis to name a few.[16]

Cardiovascular disease is a broad category for a range of diseases that are affecting heart and blood vessels. The early methods of forecasting the cardiovascular diseases helped in making decisions about the changes to have occurred in high-risk patients which resulted in the reduction of their risks.[17] The health care industry contains lots of medical data, therefore machine learning algorithms are required to make decisions effectively in the prediction of heart diseases.[18]

Identifying the processing of raw healthcare data of heart information will help in the long term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment toward heart disease.[19] Heart disease prediction is challenging and very important in the medical field.[20]

A need to develop such a medical diagnosis system situation day by day. The important key points of such medical diagnosis systems are reducing cost and obtaining more accurate rate efficiently.[21] Developing a medical diagnosis system based on machine learning for prediction of heart disease provides more accurate diagnosis than traditional way and reduces cost of treatment.[22]

Data mining is extracting information and knowledge from huge amount of data. Data mining is an essential step in discovering knowledge from databases. There are numbers of

databases, data marts, data warehouses all over the world. Data Mining is mainly used to extract the hidden information from a large amount of database. Data mining is also called as Knowledge Discovery Database (KDD).[23]

The main feature of Data Mining includes classification, mapping, and clustering. Due to the growing demand, different data mining techniques are used for better decision-making in the field of medicines too.[24]

Many medical organizations face a major challenge with quality services such as the correct diagnosis of patients and the provision of treatments where they occur often the man can pay the costs for their cure. Data mining techniques simplify many important and critical aspects problems related to health. In India and all abroad countries Heart Disease is a major challenge of medical science and specially it becomes dangerous due to lack of its prediction henceforth patient's survival becomes compromised cause death.In the year of 2003 nearly 17.3 million people died around the world and out of this, approximately 9 million were only because of the coronary heart disease. In 2008 17.3 million people died due to heart disease. More than 80% of passing away in world is because of coronary illness. WHO estimated by 2030 almost 23.6 million persons will pass away due to heart disease.[25]

There are many such factors such as smoking, alcohol, obesity, high blood pressure, diabetes etc which creates such disease and responsible for the risk of having a heart problem. However nowadays, we can avoid such kind of diseases by getting better decisions at early stage with the help of advance techniques. [25]The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors a new prominent sector is healthcare. The healthcare environment is still information rich" but knowledge poor". There is a wealth of data available within the healthcare systems. Researchers are using data mining techniques for the diagnosis of many diseases such as heart disease, diabetes, stroke and cancer and many data mining techniques together with machine learning algorithms have been used in the diagnosis of heart disease with good accuracy.[26]

We propose a new and different approach to mine frequent patterns as discriminative features. It builds a Hierarchical structure that sorts or partitions the data onto nodes from the whole list. Then at each node, it directly discovers a discriminative pattern to further divide its examples into purer subsets that previously chosen patterns during the same run cannot separate. Since the number of examples towards leaf level is relatively small, the new approach is able to examine patterns with extremely low global support that could not be enumerated on the whole dataset by the batch method as given in Fig, so in this paper had combined data mining techniques with ACO for better heart disease prediction. In this paper we use data mining to emphasize to discover knowledge that is not only accurate, but also comprehensible for the users.[27]

The data mining has four main techniques namely Classification, Clustering, Regression, and Association rule. Data mining techniques have the ability to rapidly mine vast amount of data. Data mining is mainly needed in many fields to extract useful information from a large amount of data. The fields like the medical field, business field, and educational field have a vast amount of data, thus these fields data can be mined through those techniques more useful information. [28]

Data mining techniques can be implemented through a machine learning algorithm. In this system, a heart disease data set is used. The main aim of this system is to predict the possibilities of occurring heart disease of the patients in terms of percentage. This is performed through data mining classification techniques. Classification technique is applied to the dataset through the machine learning classification algorithm. These algorithms are used to enhance the accuracy level of the classification technique. This model performs both the classification and prediction methods. These models are performed using python Programming Language.[29]

The performance of the diagnosis model is obtained by using methods like classification, accuracy, sensitivity and specificity analysis.[30]

In the heart disease prediction system, there are input variables, which are disease risk factors which are obtained from dataset, and output variables, which are a category, such as “disease absence” and “disease presence”. Prediction of heart disease is called supervised

learning problem. Because of having output variables are in category type, the prediction heart disease is “classification type of supervised learning”.

1.4Motivation:There are tremendous changes happening to our daily routines which make an impact on our health positively and negatively.Heart disease is the leading cause.It is a serious disease since we often hear that most of the people die out of Heart diseases.We have an opportunity to provide early diagnoses and treatments for people who are likely to have heart failure and help them have longer, more active lives.

1.5 Problem Definition: As there isan increase in heart diseases day by day. Many people are facing problems related to heart. People do die due to heart diseases. So we are going to observe and predict whether the person will have the heart disease or not.

1.6 Objectives:

1.Collection, preparation, cleaning and processing of dataset.

2.Feature extraction from dataset.

3. Formulation of algorithm for classification.

4. Experimentation and validation of results.

CHAPTER 2

LITERATURE SURVEY

An increasing number of heart patients worldwide have motivated researchers to do comprehensive research to reveal hidden patterns in clinical datasets. An overview of reported computational studies on pattern recognition in heart disease is covered in this section. Not only are different techniques addressed, but also various heart disease datasets are provided. Finally, the gap in existing literature, which was the main motivation of this study is also mentioned. Some of the key studies are as follows:

Santhana KrishnanJ, et al.has givendetailed information about Coronary Heart diseases such as its Facts, Common Types, and Risk Factors has been explained in this paper. The Data Mining tool used is WEKA (Waikato Environment for Knowledge Analysis), a good Data Mining Tool for Bioinformatics Fields. The all three available Interface in WEKA is used here. Naive Bayes, Artificial Neural Networks and Decision Tree (J48) are Main Data Mining Techniques and through this techniques heart disease is predicted in this System. [1]

Dr. B. Umadevi and M. Snehapriyavarious approaches of heart attack disease prediction research papers are analyzed and studied. The prediction accuracy of existing systems can be improved, so in future, new algorithms and techniques are to be developed which overcome the drawbacks of the existing system. Medical Data Mining is a domain of challenge which involves a lot of misdiagnosis and uncertainty. [2]

Sana Bharti, et al various algorithms that can be used to predict the heart disease and compare them to find out the best method of prediction of disease.combing these algorithms with various data mining techniques like clustering, classification, association rule, or hybrid these algorithms with each other will lead to better performance and high accuracy rate.[3]

AditiGavhane and Isha Pandya. et al.,the main causes or the factors which have strong influence on the heart health. Some factors are unmodifiable like age, sex and family

background but there are some parameters like blood pressure, heart rate etc. MLP provides its users with a prediction result that gives the state of a user leading to CAD. [4]

TülayKarayÕlan, et al., Cleveland database was used for heart disease prediction system. Because Cleveland database is the most commonly used database by ML researchers. neural network in the system used 13 clinical data which are obtained from Cleveland Dataset as input. [5]

CincyRaju and Philipsy E. et al., there are a lot of studies on prediction of heart disease. Results of these studies vary up to almost accuracy of 100%.The proposed system gives 95% accuracy rate which means a very good rate according to related studies on this field.techniques for the prediction of heart disease: decision tree, Support vector machine (SVM), Neural network, K-nearest neighbor algorithm. [6]

SENTHILKUMAR MOHAN et al., various methods have been used for knowledge abstraction by using known methods of data mining for prediction of heart disease.Neural networks are generally regarded as the best tool for prediction of diseases like heart disease.The results of the methods are 82.18%, 85.82% and 91.30% of UCI data sets of Cleveland. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. [7]

BO JIN and CHAO CHE et al., prediction and earlier detection of heart failure could lead to improved outcomes through patient engagement and more assertive treatment. Previous work on the early detection of heart failure has relied on conventional modeling techniques such as logistic regression(LR)and support vector machines (SVM), using features that represent the aggregation of events in an observation window and exclude the temporal relations among events in the observation window. [8]

Mamatha Alex P and Shaicy P Shaji.disease Forecasting System Using Data Mining Methods Heart disease is a term. Which is coined to describe the big quantity of health cares that are related to the heart. The defines of medicinal state the uncertain health conditions.Diagnosis of Heart Disease Patients Using Fuzzy Classification Technique The heart disease prediction is primary concern in medical science. [9]

Dinesh Kumar G., Santhosh Kumar D., et al., data which is reconconceivable outcomes of anticipating the hazard levels of patients for a colossal informational collection and sending the application on cloud stage where specialists and patients can sign in with a one of a kind ID made by them. It can be utilized by the social insurance directors to show signs of improvement administrations. Coronary illness was the most significant explanation behind casualties.[10]

Archana L. Rane., et al., designed taxonomy based on number of techniques used for HDP. different technologies generate different accuracy which also depends on data size, number of attributes considered and tools used for implementation.HDP are divided into two categories based on number of techniques used for prediction, namely "Discrete" and "Integrated". Usage of only one technique by the author in their prediction model is categorized as discrete whereas combination of multiple techniques is categorized as integrated. [11]

T.Nagamani and S.Logeswari.backpropagation algorithm for learning and testing neural network. The neural network weights were weighted with the help of optimization technique called genetic algorithm.multi layered network had twelve input nodes, two output nodes and ten hidden nodes.The attribute nun is the heart diseasediagnosis attribute. It was classified as presence and absence. If it is presence, then value of nun would be low or medium or high or very high. If it is absence, then value of nun would be zero. [12]

Jan Bohacik, et al., performance of the algorithm presented in Section III are described here and it is compared with other well-known data mining techniques as well. The algorithm consisting of Naive Bayes, discretization and validation tools is implemented in the Java programming language.The discretization method took initial divisions of values from experts and used them to produce the final discretization of numerical attributes. [13]

Salma Banu N.K.et al., different techniques used by researchers in detection of HD. Enormous technologies of DM are involved in design of HD prediction model. The Main objective is to identify the key patterns and features from the medical data of the patient by combining data mining techniques along with big data analytics to predict the heart disease before it causes to help the medical practitioners. [14]

Monika Gandhi, et al., data mining techniques brings with set of techniques to find out hidden patterns for making decision in healthcare organizations.Classification methods of data mining used in data discovery. Different classification techniques of data mining have merits and demerits for data classification and knowledge extraction. [15]

Rashmi G Saboji, etal., predict the diagnosis of heart disease with reduced number of attributes. Each dataset stored in HDFS is classified based on attributes. Thirteen attributes and one class are chosen for the prediction. This prediction solution using random forest on Apache Spark gives massive opportunity for health care analysts to deploy this solution on ever changing, scalable big data landscape for insightful decision making. [16]

Theresa Princy. R. et al., the Naive Bayes classifier algorithm uses conditional independence; it believes that an attribute value of a given class is independent of the values of other attributes. Web based health care detection was proposed. Various Data mining techniques and classifiers are discussed in many studies which are used for efficient and efficacious heart disease diagnosis. [17]

Rifki Wijaya.et al., system to analyze existing heart disease is already done, but to predict of one year ahead with all tools is not yet done. These tools are tools that are used daily by many people. Most people personally have these tools personally. Prediction is done by inserting a few variables that be a symptom of heart disease. Predictions generally only use the initial medical record alone.[18]

PiseSatishPrakashrao et al., the results of our experimental analysis in finding significant patterns for heart attack prediction are presented in the above section. With the help of the database, the patterns significant to the heart attack prediction are extracted using the approach discussed. We proposed an efficient concept for detection of Heart Diseases based on ant colony optimization and data mining. We use random ant generation for threshold setting and on the basis of that we find the global optimum value for each spectrum, so the detection is easy.[19]

P. Priyanga1 et al., diagnosis of heart disease by using machine learning methods is one of the challenges in the health field.Various Data mining techniques and classifiers are discussed in many studies which are used for efficient heart disease prediction. All the

techniques discussed could prove valuable in helping to address some of the challenges associated with reducing the expected healthcare spending due to CVD by informing, engaging and empowering individuals to actively participate in modifying their most significant risk factors for CVD. Hence, understanding the usefulness of data mining for early diagnosis of various heart diseases becomes important.[20]

Kusuma.S et al.,heart disease is one of the global health challenges in recent years. At present, many research works were carried out to predict and diagnose the heart diseases. In this study, a systematic effort was made to identify, and review machine learning, data mining and deep learning approaches applied on HD research. EHR 's in health care producing huge amount of data with the development of technology also give rise to in-depth exploration towards accurate disease prediction, diagnosis, and treatment.[21]

Amandeep Kaur and Jyoti Aroradata mining is the way toward extracting interesting patterns and knowledge from huge amount of information. The Data Mining process is a combination of choosing, analyzing, planning, interpreting and evaluating the outcomes [1]. Numerous clinical finding achievement in the data mining techniques for prediction and clustering.[22]

Nimai Chand Das Adhikariin this section we have presented the results comparison for die rent algorithms that we used in the model. Here, we have analyzed the details of 1094 patients having label as 1 or 0. Here 1 is represented for the patient's souring from any kinds of heart disease and vice versa. Also, for small plaques, the label is given as 0. For training and validation to check how the algorithm is performing, we have used the holdout technique with 70:30 ratios. There are many others cross validation techniques but we have fixed our model to start the testing phase with the 70:30 percent holdout technique.[23]

AnimeshHazra, Subrata Kumar Mandal el at.,heart diseases when aggravated spiral way beyond control. Heart diseases are complicated and take away lots of lives every year.When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time. Sedentary lifestyle and excessive stress in today's world have worsened the situation. If the disease is detected early then it can be kept under control.[24]

Susmitha K, B. Senthil Kumarheart disease is one of the major problems in nowadays which leads to causality. Predicting heart diseases is possible only by the consideration of attributes. This analyzing method of the attributes can be achieved by the inclusion of data mining techniques. Data mining methodologies embraces methods such as Neural networks, naïve bays, clustering mechanisms, classification, big data, etc. Further implementation has to be done in order to predict heart disease in a big data environment.[25]

K.Polaraju, D.Durga Prasad Today diagnosing patients correctly and administering effective treatments have become quite a challenge. Most hospitals today use some sort of hospital information systems to manage their patient data in the form of numbers, text, charts and images. The diagnosis of diseases is a critical and complicated job in medicine. The cost to treat a patient with a heart problem is quite high and not affordable by every patient. The recognition of heart disease from diverse signs is a mysterious problem that is encountered with number of false assumptions and is frequently accompanied by impulsive effects.[26]

DivyaAnnepu, Gowtham Gfrom the above work, it is observed that the implementation of ML based techniques for heart disease identification is improving the accuracy and reducing the cost factor. Almost, all the work identifies the possibility of heart failure without any major medical infrastructure equipment but with intelligent ML techniques. Using random forest, we built a platform which can be used to classify the heart patients with an accuracy of 97.5%. In future, an interactive platform can be built between the doctors and the patients to communicate about risks and factors. [27]

Fahd Saleh Alotaibithe ratio of heart failure patients has been increasing every day. To overcome this dangerous situation and deteriorate the chances of heart failure disease, there is a need of a system that can generate rules or classify the data using machine learning approaches. Therefore, this research discussed, proposed and implemented a machine learning model by combining five different algorithms. [28]

Uma N Dulhareit concluded that the proposed model is effective and efficient to improve the accuracy of the Naive Bayes classifier using the particle swarm optimization for feature

subset selection which achieves similar or even better classification performance. The goal was successfully achieved by developing a novel algorithm maximizing the classification performance and minimizing the number of features [29]

Do Thanh Thai, Quang Tran Minh et al., In this work, we have presented a lightweight approach to processing ECG signals from sensors in order to perform diagnosis based on a heart disease knowledge-base. Our preliminary results demonstrated that the proposed model can remove noises from raw ECG signals, extract under grant number 102.01-2016.28 and by The University of Dayton Research Council Seed Grant. [30]

LaxmiVerma el at.,literaturereviewsuggested thatmodels withthebestclassification performance maydifferfromoneproblemtoanother; they rely on data preprocessing techniques, feature selectionmethods,selectionofalgorithmsformodelconstruction and validation. [31]

Hlaudi Daniel Masethe, Mosima Anna Masethethe algorithms are applied on the data set using stratified 10-fold validation in order to assess the performance of classification techniques for predicting a class.This comparative study also recommends that the significantly evaluated classifier can be used for real-time prediction of heart disease patients and for predicting the risk factor of heart failure with a view to ensuring additional care so that early-stage heart failure can be avoided. However, more training data whether from hospitals or from domain-experts can be added for increasing the prediction performance of the classifiers. Moreover, diverse feature reduction strategies may also be applied on the dataset for getting improved performance.[32]

Md. Fazle Rabbi, Md. Palash Uddin et al., As heart disease is one of the vital causes to death, it should be correctly detected at very early stage to get recovery from it. Sometimes, real-life practitioner may not be able to detect the disease due to some lack of skilled knowledge and proper experiences. Thus, computer-based competently accurate prediction system may be an alternative to detect the heart disease for fixing it immediately. Hence, in this paper, three mostly used data mining classification techniques such as SVM, KNN and ANN have been studied and evaluated using standard Cleveland heart disease dataset. It

has been analyzed that RBF kernel based SVM can outperform KNN and ANN on the basis of the classification rate while KNN is also offering better performance than ANN. [33]

Gagandeep Kaur et al., amost extreme expensive and predominant medical issue anguished by the present age is the diverse heart ailments. Consequently, early foreseeing the sickness helps in turning away the future sufferings by slight changes in the way of life already. Then again finding of heart disease assumes a significant job to think about the status of heart patients. This work is identified with supervised learning for Heart disease prediction analysis. [34]

Mamatha Alex P and Shaicy P Shajithe main motivation of this project is to provide an insight about detecting and curing heart disease using data mining technique.From the medical profiles twenty attributes are extracted such as age, sex, blood pressure and blood sugar etc. to predict the likelihood of patient getting heart diseases. [35]

Salma Banu N.K el at., in this paper, a survey conducted from 2004 to 2015 gives the idea of different models available and the different data mining techniques used. The accuracy obtained with these models is also mentioned. It is observed that all the techniques available have not used big data analytics. Use of big data analytics along with data mining will give promising results to get the best accuracy in designing the prediction model. [36]

T. Nagamani, S. Logeswari, B.Gomathythis study uses Map reduce algorithm by comparing meta-heuristic approach along trained persistent fuzzy neural network on UCI machine learning repository dataset for predicting heart disease. It has some latency due to batch processing, so in future batch processing is reduced then we can get more accurate output on comparing with other data mining techniques. [37]

Dinesh Kumar G el at., this paper contributes the correlative application and analysis of distinct machine learning algorithms in the R software which gives an immediate mechanism for the user to use the machine learning algorithms in R software for forecasting the cardiovascular diseases. [38]

Michal Zabovsky el at., a decision support tool for automatic diagnosis of heart disease with a Naive Bayes classifier using a supervised discretization which takes domain expert knowledge into consideration was presented. The discretization method took initial

divisions of values from experts and used them to produce the final discretization of numerical attributes. [39]

Archana L. Rane increase in heart patients each year is a constant intimidation to each individual and a recurring problem for health authorities. Forecasting heart disease based on various attributes of a patient can help authorities take measures to handle unexpected situation. [40]

Hence, From the above research papers we have studied and we have observed the different methods, concepts, causes, heart disease types, and how the prediction of the heart attack can be done by using various algorithms and methods. Therefore, we decided to use the Machine Learning Language.

CHAPTER 3

BLOCK DIAGRAM / DESCRIPTION/ FLOW CHART ALGORITHM

Now we are going to discuss about the block diagram of our project where we will see which algorithms we are going to use.

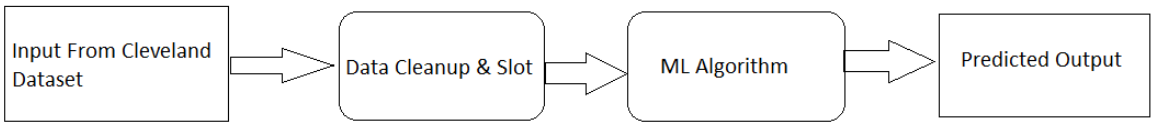


Fig 3.1. Block Diagram of Prediction the Risk of Heart Attack.

Description:

Database-

- Cleveland database was used for heart disease prediction system. Because Cleveland database is the most commonly used database by ML researchers. The dataset contains 303 instances and 76 attributes,but only 14 of them are referred by all published studies. The "goal" field which has varying values from 0(absence) to 4 denotes if heart disease present or not in the patient. Studies on the Cleveland database have focuses on distinguishing absence(value 0) from presence (values range from 1 to 4) [13]. The dataset has some missing values in it. Firstly, missing values were filled with interpolation values. Then dataset was split into three parts: one for training (%70), second one for testing (%15) and third one for validation(%15). There are 213 instances and 13 attributes in training data. Test data and validation data contain 45 instances and 13 attributes.

- A Cleveland dataset is a collection of information that is organized so that it can be easily accessed, managed and updated. Dataset process workloads to create and update themselves, querying the data they contain and running applications against it.
- In this project of Prediction of the Risk of Heart Attack after finding different parameters that could help analyze the person's risk of having CVD. A huge database is required to store the information.
- Mentioned dataset is taken from different patients considering their age, gender, blood pressure, family history, cholesterol.

Advantages: -

- The Datasets will help in to be more precise in analyzing risk of CVD.
- The dataset is available and standardize.
- It is easier to enforce as all data in database can be accessed through centralized DBMS.
- Integrity and flexibility can be added.

Disadvantages: -

- Huge dataset causes increase in complexity.
- Substantial cost and time required in training.

Data Set cleanup & sort out:

The database of patients is collected and the most accurate and major attributes and feature which help in predicting heart disease such as: age(in years),sex(m/f),chest pain type(typical angina, atypical angina, non-angina, asymptomatic),resting blood pressure(in mmHg),cholesterol(mg/dl), high fasting blood sugar(1/0),resting electrocardiographic results(1/0), maximum heart rate achieved, resting heart rate , exercise induced angina(1/0),

St depression induced by exercise relative to rest,number of color vessels, obesity, thal(3=normal, 6= fixed defect, 7=reversible defect) are considered.

ML Algorithm:

Various algorithms can be applied here to explore the best attribute which will evaluate the fitness value which is assigned to each attribute(individual). Algorithms which can be used can be PSO, Genetic Algorithm, Ant Colony Optimization, and Artificial Neural Network or may be the hybrid of any of these algorithms.

Predicted output:

Predicts whether the person has a heart disease or will suffer in future.

FlowChart:

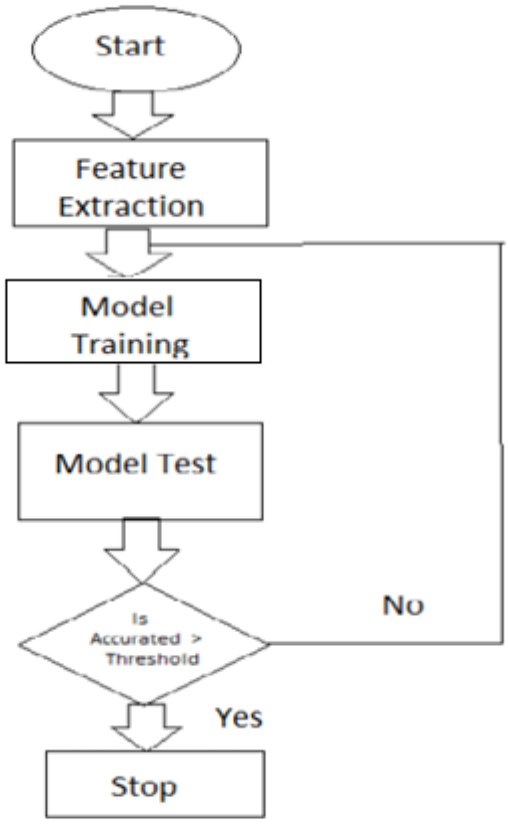


Fig.3.2 Flow Chart

Algorithm:

1] Start

2] Collect required information from dataset.

3] Set model for prediction

4] Test the model

5] If data is not correct then go to step 2

6] If data is correct then give predicted output

7] Stop.

Predicted Flow Chart:

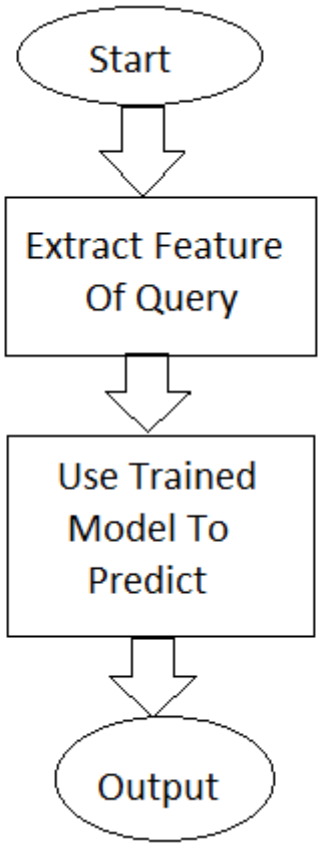


Fig.3.3 Predicted Flow Chart

Algorithm:

1] Start

2] Find out the Query from dataset & solve it.

3] Used trained model for prediction

4] Predicted output.

CHAPTER 4

METHODOLOGY

The Methods that are going to be used and are being implemented are discussed. The below figure shows the different processes. We have identified and applied the various techniques which help for observation of the prediction of the heart disease. The detail description is given below.

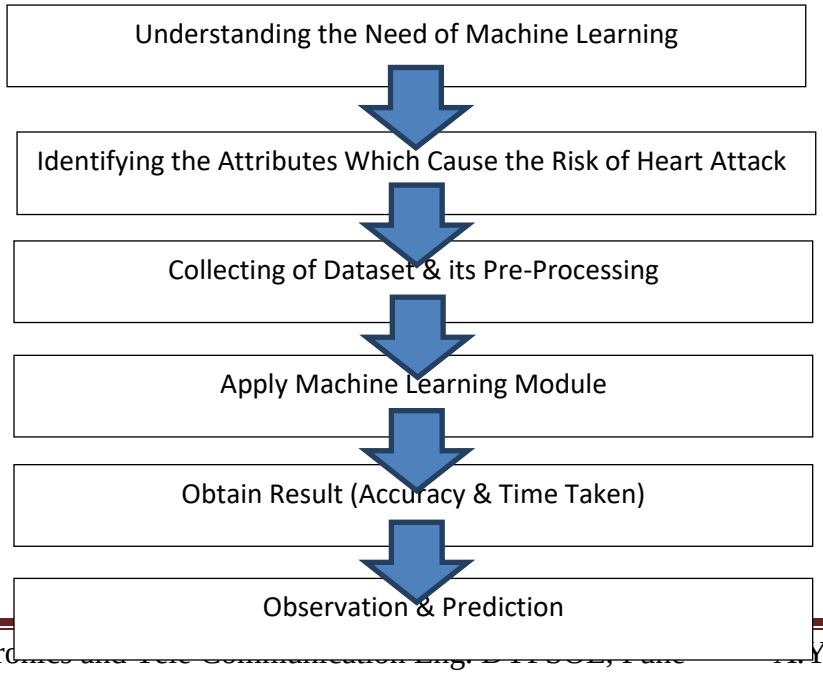


Fig.4.1 Proposed Methodology

1.Machine Learning: Machine Learning is extracting information and knowledge from huge amount of data. Machine Learning is an essential step in discovering knowledge from databases. There are numbers of databases, data marts, data warehouses all over the world. Machine Learning is mainly used to extract the hidden information from a large amount of database.

2.Pre-processing: The pre-processing is the first stage of the heart disease prediction research process. In this phase, the dataset is taken as input and data set is cleared by discarding missing qualities.

3.Highlight Extraction: In the second stage, the features of the input data are extracted for the classification and the connection between the different attributes is buildup for the recognizable proof of main attributes from the dataset. The strategy of decision tree classifier is applied in this stage for the feature extraction.

4. Grouping: In the last stage, the relationship between the attributes is taken as input for this stage, the input data is divided into training set and test set. The algorithm of decision tree will be applied on the extracted features for the final classification. The decision tree will behave as base classifier for the classification, which is the classification approach used for prediction. The execution of proposed approach will be investigated as far as specific parameters like exactness, accuracy, review and f-measure. Cleveland Clinic Foundation dataset known as "Cleveland Clinic Foundation Heart Disease Dataset" utilized for the examination. Dataset include 14 attributes in which 303 samples as a patient's data were consider in it from which 13 attributes such as age, sex, chest pain, serum-cholesterol, fasting blood sugar, electrocardiographic, heart rate, induced angina, old peak, slope and thal are taken as input parameters and last 1 attribute define the diagnosis of Heart disease. Specific data mining techniques were applied on the dataset by using cross validation method to validate the results. This technique classified the data into two parts such as training set and testing set from which the dataset with small number of samples were consider in testing dataset and data set with large number of sample were consider in

training dataset. The benefit of contribution to weight transformation is that it decreases the quantity of correlations between the client esteems and the records of the dataset.

CHAPTER 5
RESULT AND DISCUSSION

The aim of this project is to know whether the patient has heart attack or not. The record in the datasets are divided into training set and test sets. After preprocessing the data, data were applied. This section shows the result of those classification model done by python programming and then the result is generated.

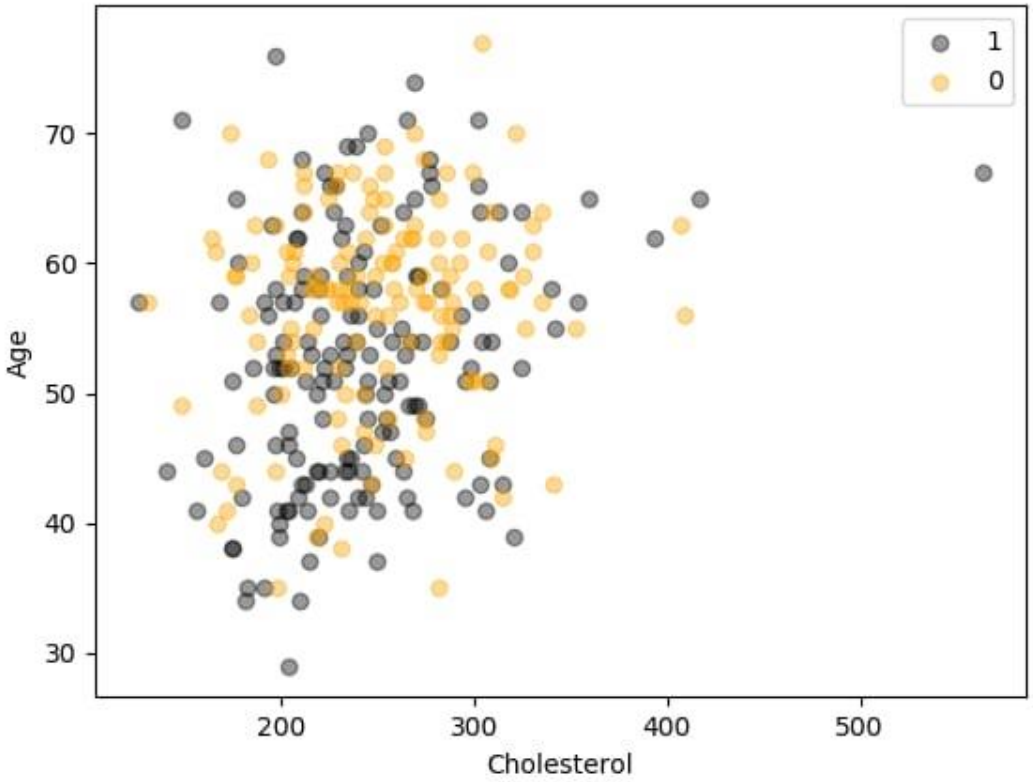


Fig 5.1 Cholesterol Vs. Age

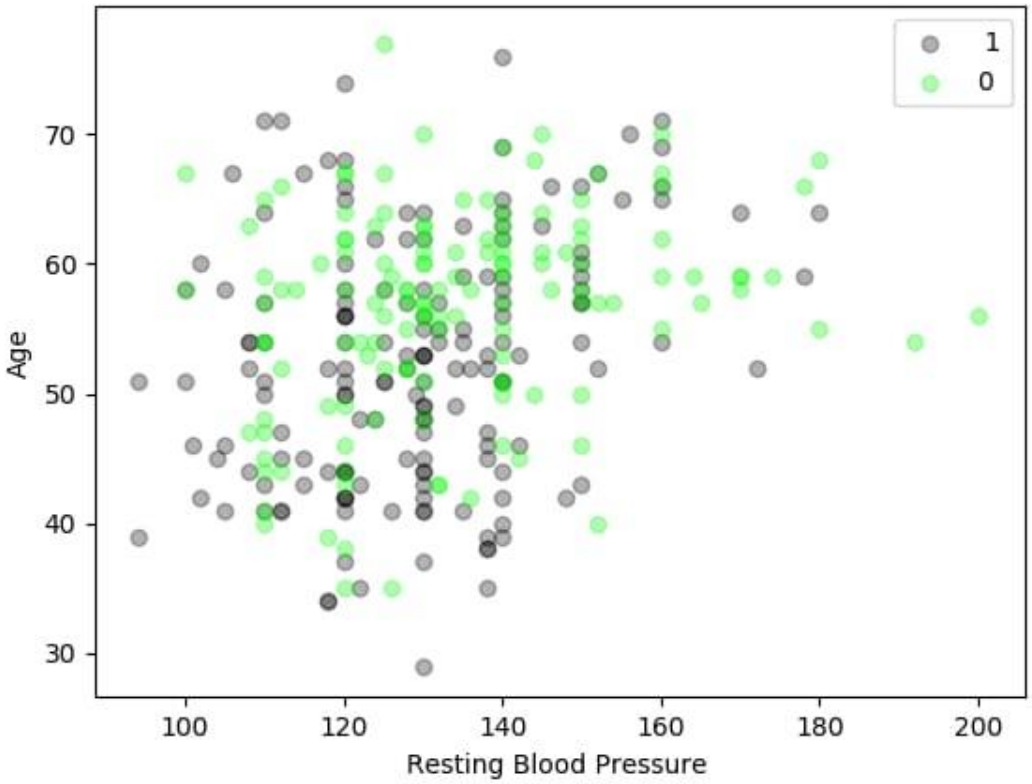


Fig 5.2 Resting Blood Pressure Vs. Age

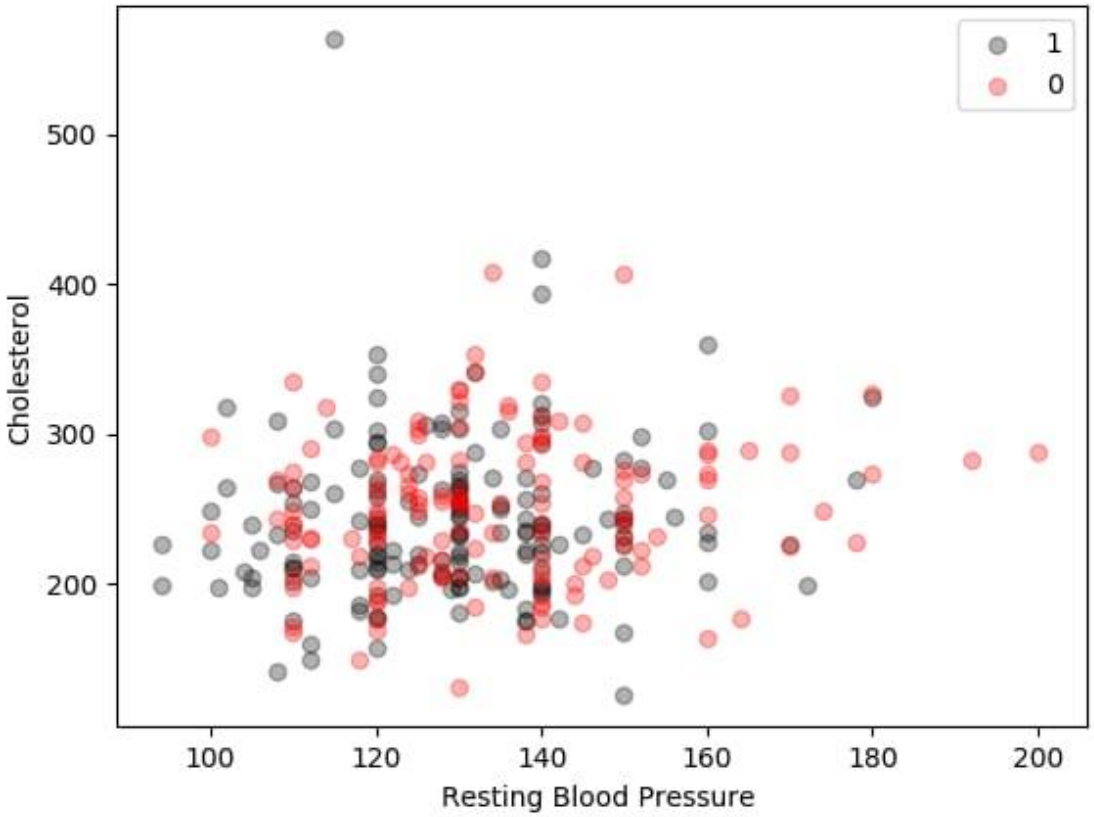


Fig 5.3 Resting Blood Pressure Vs. Cholesterol

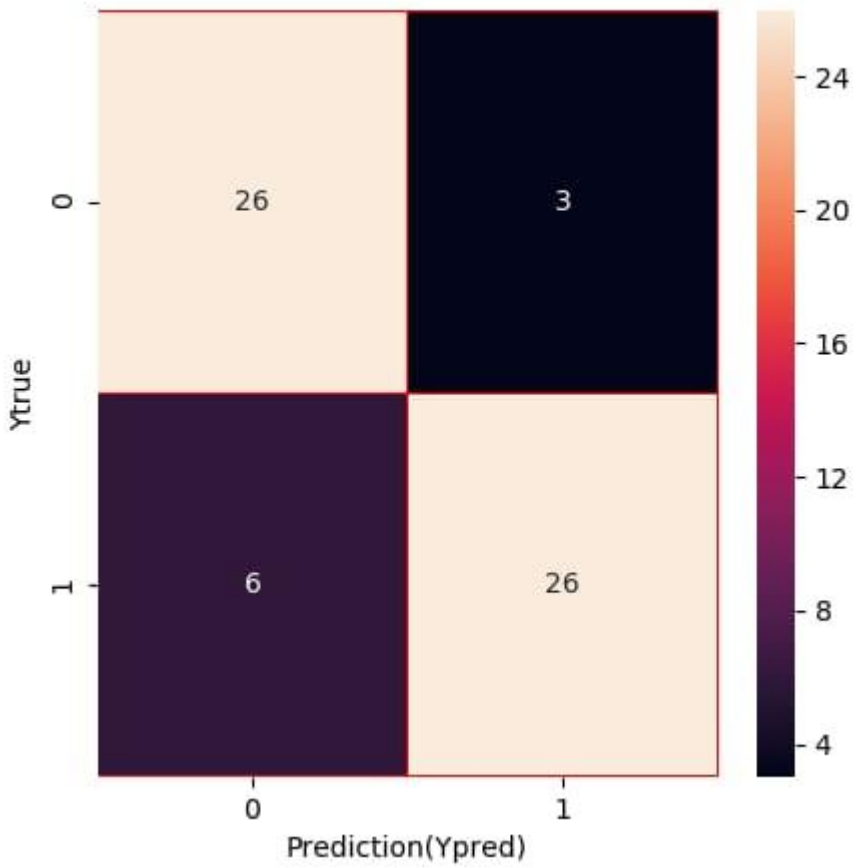
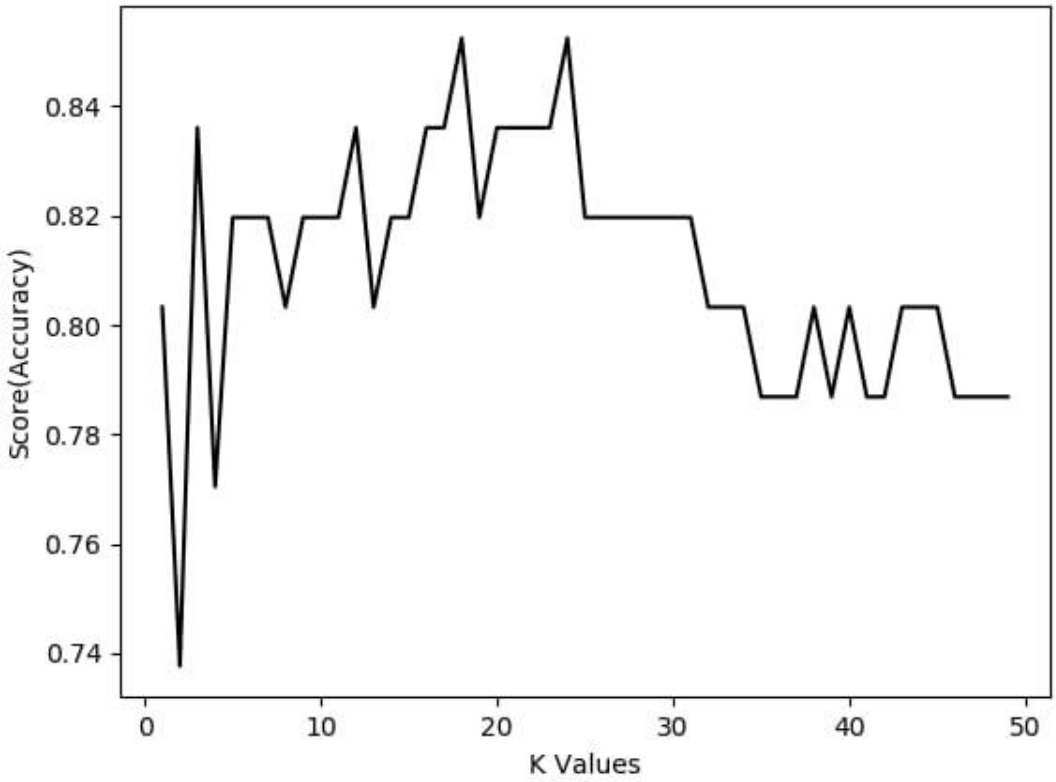


Fig 5.4 Main Regression


0.84
0.82
0.80
0.78
0.76
0.74
Score(Accuracy)
0
10
20
30
40
50
K Values

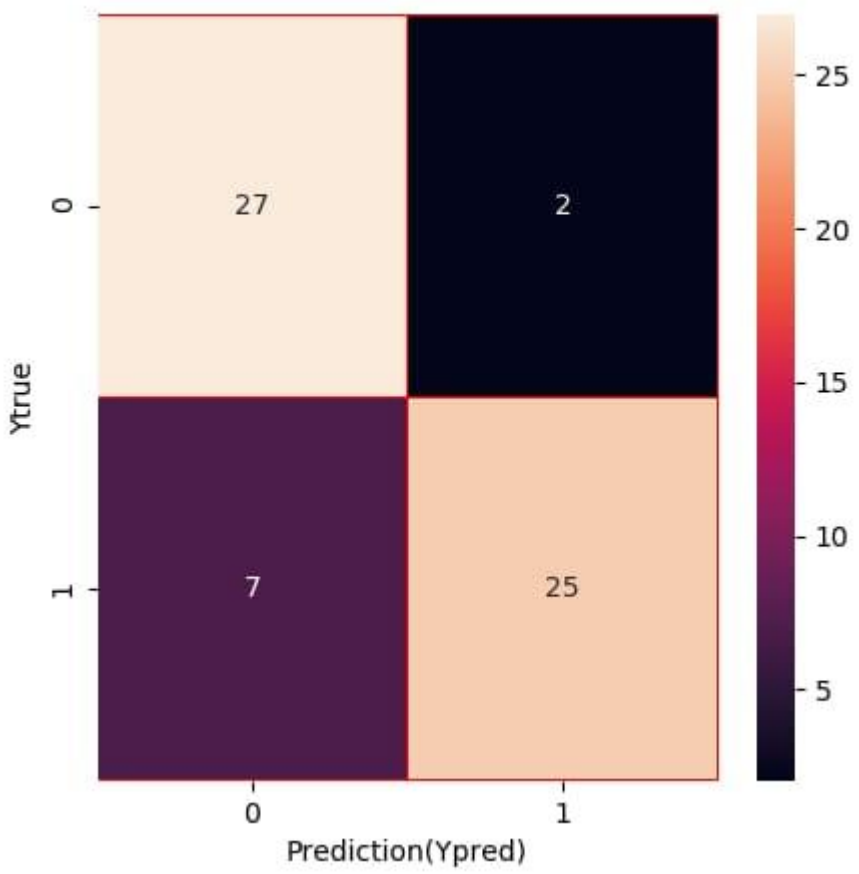
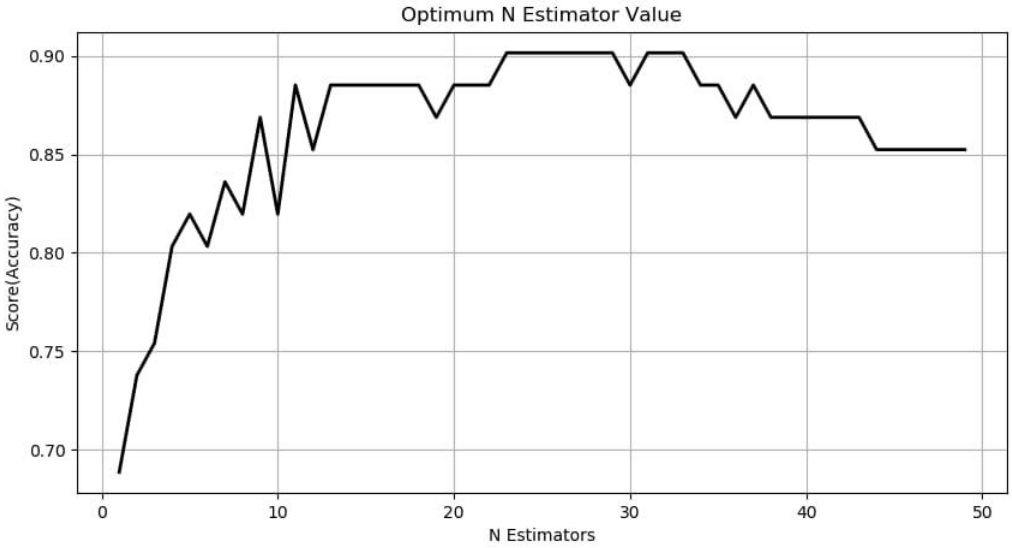


Fig 5.5 Main KNN



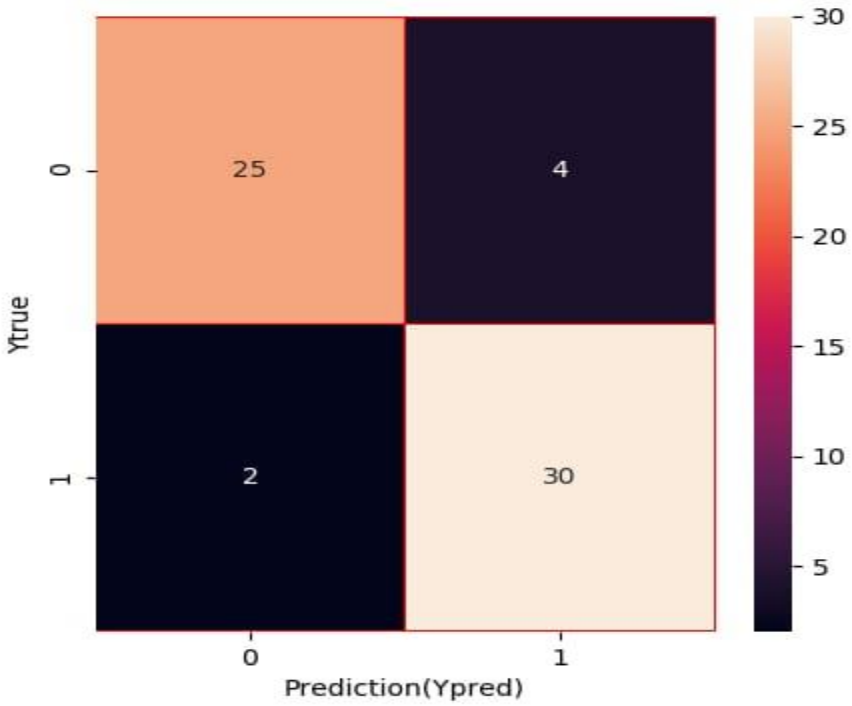


Fig no. 5.6 Random forest

REFERENCES

[1]S. K. J and G. S, "Prediction of Heart Disease Using Machine Learning Algorithms.," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019.

[2]Dr.B.U and M.S "A System for Heart Disease Prediction Using Data Mining Techniques," International Journal of Advance Engineering and Research Development, vol. 2, no. 10, 2015.

[3] S. Bharti and S. N. Singh, "Analytical study of heart disease prediction comparing with different algorithms," International Conference on Computing, Communication & Automation, 2015.

[4]AditiGavhane and Isha Pandya, "Prediction of Heart Disease Using Machine Learning," 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)

[5]T. Karayilan and O. Kilic, "Prediction of heart disease using neural network," 2017 International Conference on Computer Science and Engineering (UBMK), 2017.

[6]CincyRaju and Philipsy E. et al., "A Survey on Predicting Heart Disease using Data Mining Techniques"Proc. IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018) 2-3 March 2018, Mahendra Engineering College, Tamilnadu, India.

[7]S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019.

[8] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the Risk of Heart Failure with EHR Sequential Data Modeling," IEEE Access, vol. 6, pp. 9256–9261, 2018.

[9] M. A. P and S. P. Shaji, "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique," 2019 International Conference on Communication and Signal Processing (ICCSP), 2019.

[10]K. G. Dinesh, K. Arumugaraj, K. D. Santhosh, and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), 2018.

[11] A. L. Rane, "A survey on Intelligent Data Mining Techniques used in Heart Disease Prediction," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018.

[12]T. Nagamani, S. Logeswari, et al., "Heart Disease Prediction using Data Mining with Mapreduce Algorithm," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8,2019

[13]J. Bohacik and M. Zabovsky, "Naive Bayes for statlog heart database with consideration of data specifics," 2017 IEEE 14th International Scientific Conference on Informatics, 2017.

[14]N. K. S. Banu and S. Swamy, "Prediction of heart disease at early stage using data mining and big data analytics: A survey," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2016.

[15]Monika Gandhi, et al., "Predictions in Heart Disease Using Techniques of Data Mining,"2015 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management

[16] S. Babu, E. M. Vivek, K. P. Famina, K. Fida, P. Aswathi, M. Shanid, and M. Hena, "Heart disease diagnosis using data mining technique," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017.

[17]N. K. S. Banu and S. Swamy, "Prediction of heart disease at early stage using data mining and big data analytics: A survey," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2016.

[18]R. G. Saboji, "A scalable solution for heart disease prediction using classification mining technique," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017.

[19]PiseSatishPrakashrao"An efficient approach for detection of heart attack using Nobel and colony optimization concept of data mining,"International journal of engineering science & research technology (IJESRT)2018.

[20]P. Priyanga1"A Literature Survey on Mining Health Data to Predict Heart Disease," International Journal of Innovative Research in Computer and Communication Engineering 2017.

[21]Kusuma.S "Machine Learning and Deep Learning Methods in Heart Disease (HD) Research," International Journal of Pure and Applied Mathematics 2018.

[22]Jyoti Arora "Heart disease prediction using data mining technique: A survey,"International Journal of Advanced Research in Computer Science 2018.

[23]Nimai Chand Das Adhikari "prevention of heart problem using artificial intelligence,"International Journal of Artificial Intelligence and Applications (IJAIA), Vol.9, No.2, March 2018

[24]AnimeshHazra, Subrata Kumar Mandal "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review,"Advances in Computational Sciences and Technology 2017.

[25]Susmitha K, B. Senthil Kumar "A Survey on Data Mining Techniques for Prediction of Heart Diseases,"IOSR Journal of Engineering (IOSRJEN) 2018.

[26]K. Polaraju, D.Durga Prasad "Prediction of Heart Disease using Multiple Linear Regression Model," IJEDR 2017.

[27]DivyaAnnepu, Gowtham G "Cardiovascular Disease Prediction Using Machine Learning Techniques,"International Research Journal of Engineering and Technology (IRJET) 2019

[28]Fahd Saleh Alotaibi "Implementation of Machine Learning Model to Predict Heart Failure Disease,"(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019

[29]Uma N Dulhare", Prediction system for heart disease using Naive Bayes and particle swarm optimization."Biomedical Research 2018.

[30]Do Thanh Thai, Quang Tran Minh "Toward anIoT-based Expert System for Heart Disease Diagnosis,"MAICS2017

[31]LuxmiVerma "An intelligent noninvasive model for coronary artery disease detection,"OriginalArticle 2017.

[32]Hlaudi Daniel Masethe, Mosima Anna Masethe "Prediction of Heart Disease using Classification Algorithms,"Proceedings of the World Congress on Engineering and Computer Science 2014.

[33]Md. Fazle Rabbi, Md. Palash Uddin "Performance Evaluation of Data Mining Classification Techniques for Heart Disease Prediction," American Journal of Engineering Research (AJER) 2018.

[34] Gagandeep Kaur "Heart Disease Prediction using KNN classification approach,"International Journal of Computer Sciences and Engineering 2019.

[35]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, April 4-6, 2019, India

[36] Salma Banu N.K "Prediction of Heart Disease at early stage using Data Mining and Big Data Analytics: A Survey,"2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)

[37] T.Nagamani, S.Logeswari, B.Gomathy "Heart Disease Prediction using Data Mining with Mapreduce Algorithm," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019

[38] Dinesh Kumar G "Prediction of Cardiovascular Disease Using Machine Learning Algorithms,"Proceeding of 2018 IEEE International Conference on Current Trends Toward Converging Technologies, Coimbatore, India

[39] Michal Zabovsky "Naive Bayes for Statlog Heart Database with Consideration of Data Specifics,"2017 IEEE 14th International Scientific Conference on Informatics.

[40] Archana L. Rane "A survey on Intelligent Data Mining Techniques used in Heart Disease Prediction,"3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018.