

A Review on Heart Disease Prediction using Machine Learning Algorithms

Preeti Sondhi¹, Adeen²

¹Assistant Professor, ²M. Tech Scholar

^{1,2}Universal Group of Institutions, Mohali, Punjab, India

¹preetisondhi4@gmail.com, ²adeenmir1995.am@gmail.com

ABSTRACT

Deaths from heart disease in today's era have become a major issue around one person dies every minute because of heart disease. It takes into account both male and female groups, and this ratio can differ depending on the region, and this ratio is also used for the 25-69 age group. That doesn't mean that this heart disease won't affect people with other age groups. This problem may also begin in the early age group and foresee the cause and disease is a major challenge now. Heart disease is one of the most deadly problems in the world, one that cannot be seen with a naked eye and that occurs immediately when it meets its limits. Therefore, accurate diagnosis at the exact time is required. Health care industry has generated enormous amounts of patient and illness related data each day. However, the researchers and practitioners do not make effective use of this data. Today, although low in information the healthcare industry is rich in data. There are numerous techniques and tools available in data mining and machine learning to derive useful information from databases and to use this knowledge for more accurate diagnosis and decision making. Increasing work on predicting mechanisms for heart disease, summing up the completely incomplete research on it becomes important.

Keywords: Cardiovascular Diseases, Support Vector Machines, K- Nearest Neighbour, Naive Bayes, Decision Tree, Random Forest, Ensemble Models.

INTRODUCTION

Defect may lead to sudden death. Heart disease is nowadays the major cause of death in the country. The World Health Organization (WHO) estimated that worldwide 12 million deaths occur due to heart disease every year. 17.3 million People died from Heart Disease in 2008. More than 80 percent of the world's deaths are from heart disease. WHO predicted about 23.6 million people will die from Heart disease by 2030. This is one of the reasons why researchers are more

focused on developing an intelligent system that can be used to diagnose heart disease with high precision to prevent misdiagnosis [1]. In fact, many people have no knowledge of coping with heart disease. When cardiovascular disease could be predicted sooner, more patient deaths would be avoided and a more reliable and effective treatment approach could also be established.

Life is based on heart working efficiently since heart is an essential part of our body. When heart operation is improper, it will affect other human body parts including liver, kidney, etc. Cardiac disease is a disease that affects heart surgery. Heart and blood vessel disease are also referred to as heart disease which involves numerous problems, many of which are linked to a process called atherosclerosis. Atherosclerosis is a disease that occurs in the walls of the arteries when a material called a plaque builds up. If blood clots form, blood flow can stop. This can cause a stroke or a heart attack. There are a number of factors that make heart disease more dangerous. High levels of blood cholesterol, high levels of triglycerides, high blood pressure, diabetes and pre-diabetes, overweight and obesity, smoking, lack of physical activity, unhealthy diet and tension. Heart disease is now the leading cause of death in the world, which is also confirmed by the World Health Organization report, which reports that 12 million deaths occur worldwide due to heart disease every year [2]. Half the deaths from cardio-vascular diseases occur in the US and other developed countries. It is also the leading cause of death in several developing countries. Generally, it is known to be the primary reason behind adult deaths. In the United States, heart disease kills a single person every 34 seconds. Eminent health services mean that patients are treated correctly and that care is successful. Lower medical decisions can lead to disastrous, unacceptable consequences. The health sector must also try to minimize the number of tests used to diagnose the disease. All of this can be accomplished by commissioning a suitable decision support system. Most healthcare organizations, nowadays, have used hospital information systems to handle patient data. Both programs have been designed to support the accounting, inventory management and statistical calculations. Sadly, the data is not used to make decisions. This massive data can be used to answer questions such as "predict the risk of heart disease patients." Medical diagnosis is considered an important but complicated task which must be carried out accurately and efficiently.

People undergo a routine and busy schedule in our day-to-day lives which leads to stress and anxiety. In top of that, the number of people who are obese and cigarette-addicted is rising significantly. That contributes to illnesses such as heart disease, cancer, etc. Here prediction is the problem behind these illnesses. Every person has varying pulse rate and blood pressure values. Nonetheless, scientifically established, the pulse rate must be between 60 and 100 beats

per minute and the blood pressure must be between 120/80 and 140/90. Heart disease is one of the world's leading causes of death [3]. The number of people affected by heart disease is rising in both men and women, regardless of their age but other factors such as gender, diabetes and BMI also lead to this illness. In this paper we tried to predict and evaluate heart disease by taking into account parameters such as age, gender, blood pressure, heart rate, diabetes and so on. Because multiple factors are involved in heart disease, it is difficult to predict this disorder. Some of major symptoms of heart attack are:

- Chest tightness.
- Shortness of breath.
- Nausea, Indigestion, Heartburn, or stomach pain.
- Sweating and Fatigue.
- Pressure in the upper back Pain that spreads to the arm.

The following are the heart disease forms: heart means "cardio" Hence all cardiac diseases include cardiovascular disease groups. The various types of heart disease include:

- Coronary heart diseases.
- Angina pectoris
- Congestive heart failure.
- Cardiomyopathy
- Congenital heart diseases.

The narrowing of the coronary arteries is coronary heart disease or coronary artery disease. The Coronary Arteries provide the heart with oxygen and blood. This causes a great number of people to get sick or face death. It is one of the most popular forms of heart disease.

High diabetes blood glucose can damage the blood vessels and nerves which control the heart and blood vessels. If a person has diabetes for a longer period of time, the person's chances of having heart disease in the future are high. There are other reasons for diabetes which contribute to heart disease. They smoke which increases the risk of developing cardiac disease, High blood pressure makes the heart more difficult to pump blood, and may strain the heart and damage the blood vessels, and high cholesterol levels may also lead to heart disease and obesity. Family history of cardiac disease may also be a cause of heart disease. But this history for predicting heart disease is not included in this article.

The other risk factors are age, sex, stress and poor diet. The chance of getting a heart condition rises when a person gets older. Men have an increased risk of developing heart disease. Yet after menopause, women also have the same chance. It can also damage the arteries and increase the chance of coronary heart disease by living a stressful life.

HEART DISEASE

The heart is a delicate part of the human body which according to medical experts must be carefully guarded. Heart disease can be related to any type of pain affecting the proper functioning of the human heart.

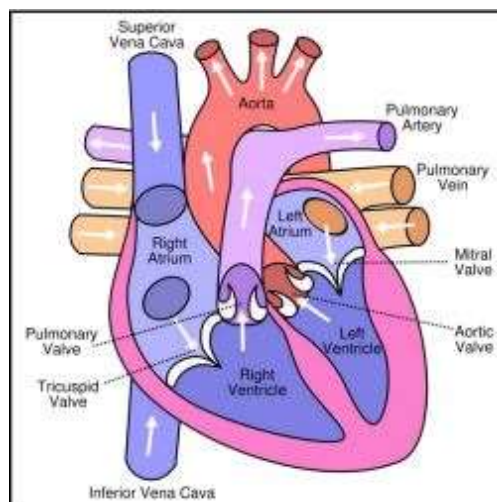


Fig1. Human Heart Structure [4]

Cardiac disease can be divided into various forms. This includes:

- (a) **Heart Attack:** Blood flow to heart is drastically reduced or stopped during heart attack. Heart attack occurs when one or more coronary arteries are blocked which carry blood to the heart. Cardiac attack can cause heart muscle damage.
- (b) **Heart failure:** This is a condition where heart fails to pump enough blood to the body that is needed. The condition does not mean that the heart has stopped functioning. Major symptoms of heart failure include rise of weight, swelling, shallow breath, cough and tiredness.

(c) **Angina:** It is an annoyance that occurs when the heart doesn't get the proper blood. Angina is a complication of coronary artery disease, causing strain, chest pain, back, hands, neck and shoulders.

(d) **Arrhythmias:** The irregular speed and slowness occurring during heartbeat.

Arrhythmias may also be called irregular beat occurring in the heart.

Table1. Heart Diseases [5]

Arrhythmia	The heart beat is unsuitable whether it is erratic, too sluggish or too fast.
Cardiac arrest	Suddenly comes an unexpected loss of heart function, consciousness and breathing.
Congestive heart failure	The heart is not pumping blood as well as it should, it is a chronic condition.
Congenital heart disease	An abnormality of the heart that occurs prior to birth.
Coronary artery disease	The major blood vessels of the heart can injure, or any disease in the blood vessels can occur.
High Blood Pressure.	It has a disorder that blood flow against the walls of the artery is too strong.
Peripheral artery disease	Circulatory disease is the narrowed blood vessels that limit blood flow in the limbs.
Stroke	Blood supply disruption damage occurs to the brain.

FRAMEWORK FOR MEDICAL DATA MINING

A general framework proposed by for medical data mining is shown in fig below.

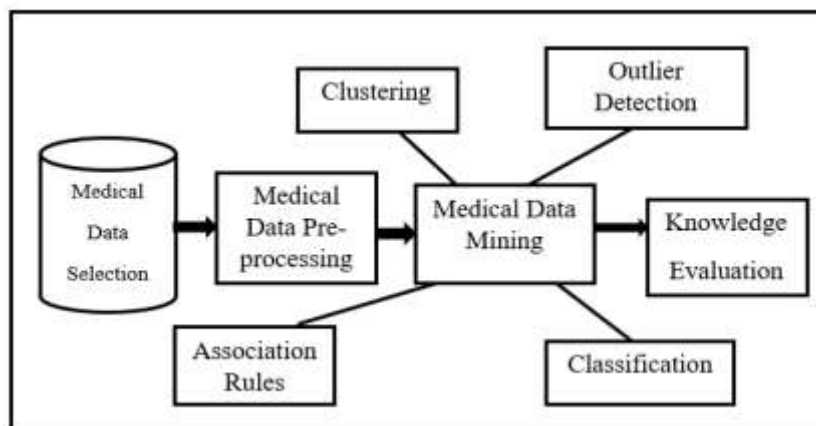


Fig 2. Framework for Medical Data Mining [11]

TECHNIQUES IN DATA MINING

Data mining techniques are tools used to organize data to identify patterns, including correlation, grouping, prediction, and clustering techniques.

- a) **Association:** It Seek trends that are focused on linking particular event to other events. It is widely used to forecast cardiac disease because it provides the relationship of unlike features for evaluating and grouping patients with all the risk factors required for prediction.
- b) **Clustering:** This is machine learning without supervision, in which no class labels are provided. It locates and visually registers sets of information that were not known beforehand.
- c) **Classification:** The method is supervised learning, where some training samples are given with the class labels. In classification methods are used mathematical methods such as linear programming, Naïve Bayes, Decision trees and artificial neural network.
- d) **Forecast(foreshadowing):** It determines data trends that can lead to a fair future prediction. The prediction methodology fits in with data mining's prognostic model.
- e) **Sequence:** Searches for patterns that lead from one event to another possible occurrence.

ALGORITHMS IN MACHINE LEARNING FOR DATA MINING

Such algorithms give computer systems the ability to learn without being programmed fully. Machine learning is closely linked to data mining, and the two approaches search for patterns through data. Machine learning, however, uses data to find patterns in data, and then correctly adjust the software instead of extracting human understanding. Data-mining machine learning algorithms:

- (a) **Supervised learning:** There is a known mark in this approach for the training data which is the input data. An algorithm is designed to adapt to new data what was learned in the past.
- (b) **Unsupervised learning:** Unknown labelling of the input data. An algorithm is built by deducting inferences from the datasets, which can be either to obtain common rules or to eliminate redundancy through mathematical process.
- (c) **Semi-supervised:** Input data includes data that is labelled and unlabelled. There is a need to predict here but the algorithm must learn the structures to properly organize the data in order to predict accurately.

LITERATURE REVIEW

BeantKaur, Williamjeet Singh [2014] Suggested analysis of the prediction method for heart disease using data mining techniques. Data mining provides the methodology and technology to turn data mounds into valuable decision-making knowledge, and to use this data with good accuracy for predicting heart disease. So that prediction provided efficient results by using data mining algorithm. The application of data mining techniques to data on the treatment of heart disease can provide as reliable performance as that obtained in diagnosing heart disease.

Sharma Purushottam et al, [2015] Proposed rules c45 and partial tree prediction methodology for heart disease. This paper will discover rules to estimate patients ' risk levels based on their safety parameter. Quality can be measured in measurements of classification of accuracy, classification of errors, rules produced and the results. Then comparison was performed using partial tree and C4.5. The result shows potential for prediction and is more effective.

Boshra Brahmi et al [2015] Through this we have developed various data mining techniques to test the heart disease prediction and diagnosis. The main goal is to compare the different techniques for the classification. Tree decision, KNN decision, SMO decision and Naïve Bayes decision. After this, several performance measures are assessed and measured in terms of accuracy, precision, sensitivity, specificity. Decision tree provides the best method for forecasting heart disease.

Noura Ajam [2015] In this we suggested diagnosis of heart disease via artificial neural network. Feed forward Back propagation learning algorithms have been used to check the model based on their ability. Classification accuracy reached 88 per cent and 20 neurons in the secret layer by considering correct feature. ANN shows results that are important for prediction of heart disease.

K. Prasanna Lakshmi, Dr. C. R .K.Reddy [2015] Proposed fast rule-based prediction of heart disease using associative mining classification. This study uses associative classification which uses Associative Classification and rewarding approach to construct a classifier with prediction rules of high interesting values and experimental results show that this work helps doctors in their diagnostic decisions.

Marjia et al, [2016] In this we developed prediction of heart disease using WEKA software with KStar, j48, SMO, and Bayes Net and Multilayer perception. Using k-fold cross validation, performance from various factor SMO and Bayes Net achieves optimum performance than KStar, multilayer vision, and J48 techniques. The accuracy performance of those algorithms is still unsatisfactory.

K. Gomathi et al, [2016] In this we proposed multi-illness prediction using data mining techniques. Data mining nowadays plays a vital role within predicting various illnesses. The number of tests can be minimized by using Data Mining techniques. This paper focuses mainly on predicting heart disease, diabetes and cancer of the breast, etc.

S.Seemaet al, [2016] Focuses on techniques capable of predicting chronic disease by using Naïve Bayes, Decision Tree, Support Vector Machine (SVM) and Artificial Neural Network(ANN) data from historical health records. In order to measure the better performance at an acceptable scale, a comparative analysis is carried out on classifiers.

Prajakta Ghadge et al, [2016] Big data suggested for predicting a heart attack. The goal of this paper is to provide prototype using the techniques of big data and data modelling. It can also be used to derive heart disease-associated trends and relationships from database. This system is made up of two databases, namely the original big dataset and one is modified. This system can help the healthcare professionals make smart decisions. Automation would be beneficial in this system.

P. Sai Chandrasekhar Reddy et al, [2017] Proposed prediction of heart disease using data mining algorithm ANN. When heart disease treatment medical costs rose, there was a need to develop new technology that can predict heart disease. Prediction model is used after assessment to predict the patient's condition on the basis of various parameters such as heart beat rate, blood pressure, cholesterol etc.

Jaymin Patel et al, [2017] Suggested heart disease prediction data mining methods and machine-learning. You can implement this program using the WEKA app. WEKA's classification instruments and explorer mode are used for testing.

Stephen F Weng et.al, [2017] Using four algorithms for machine learning, such as logistic regression, random forest, gradient boosting machines and neural networks. They have shown that the machine learning algorithms work well in correctly predicting cases of heart disease. Researchers claim this is the first application to daily patient data in electronic records using machine learning techniques. The dataset's source is Datalink Clinical Practice Research (CPRD).

Himanshu Sharma, M A Rizvi [2017] Using Decision tree, vector supporting computer, deep learning, K closest neighbour algorithms. Because the datasets contain noise, they tried to reduce noise by cleaning and pre-processing the dataset, and also attempted to reduce the data set's dimensionality. We find that the neural networks can achieve good precision.

R.Sharmila et al, [2018] suggested use of non-linear heart disease prediction classification algorithm. It is suggested to use bigdata tools like Hadoop Distributed File System (HDFS), MapReduce along with SVM to predict optimized attribute collection for heart disease. This research included an inquiry into the use of various data mining methods to predict heart disease. This recommends using HDFS to store big data in different nodes and simultaneously execute

the prediction algorithm using SVM in more than one node using SVM. SVM is used in parallel fashion that provided better processing time than sequential SVM.

Abhay Kishore et.al [2018] Our experiments showed that recurring neural networks have good accuracy in comparison with other algorithms such as CNN, Naïve Bayes and SVM. Neural networks thus perform well in predicting heart disease. We also developed a program that could detect heart attacks in silence, and warn the patient as soon as possible.

M.Nikhil Kumar et.al [2018] Various algorithms–Decision tree, random forest, Naïve Bayes, KNN, Help vector machine, logistic tree model algorithm used. Compared with other algorithms, the Naïve Bayes algorithm has given good results. They made use of heart disease dataset UCI server. Also, it took less time to develop the J48 algorithm and it produced good results.

Chala Beyene et al, [2018] Supported Cardiac Disease Detection and Evaluation Using Data Mining Techniques. The main aim is to predict the incidence of heart disease within a short time for early automated diagnosis of the disease. The approach suggested is also important in healthcare organization of professionals who have no more knowledge and skills. This uses various medical attributes such as blood sugar and heart rate, age, sex are some of the attributes included for determining whether or not the person has heart disease. Dataset analysis is performed using WEKA tools.

Adil Hussain Seh, [2019] Heart disease is one of the most fatal problems in the world, one that cannot be seen with a naked eye and that comes instantly when it reaches its limitations. Therefore, accurate diagnosis at the exact time is required. Today, however poor in knowledge the healthcare industry is rich in data. There are numerous techniques and methods available in data mining and machine learning to extract effective information from databases and to use this information for more accurate diagnosis and decision taking.

Avinash Golande, [2019] Deaths from heart disease in today's era have become a big issue around one person dies every minute because of heart disease. This takes into account both male and female categories, and this ratio can vary depending on the region, and this ratio is also considered for the 25-69 age group. This does not mean this heart disease does not impact people in other age groups. This problem may also start in the early age group and nowadays it is a major challenge to predict cause and disease.

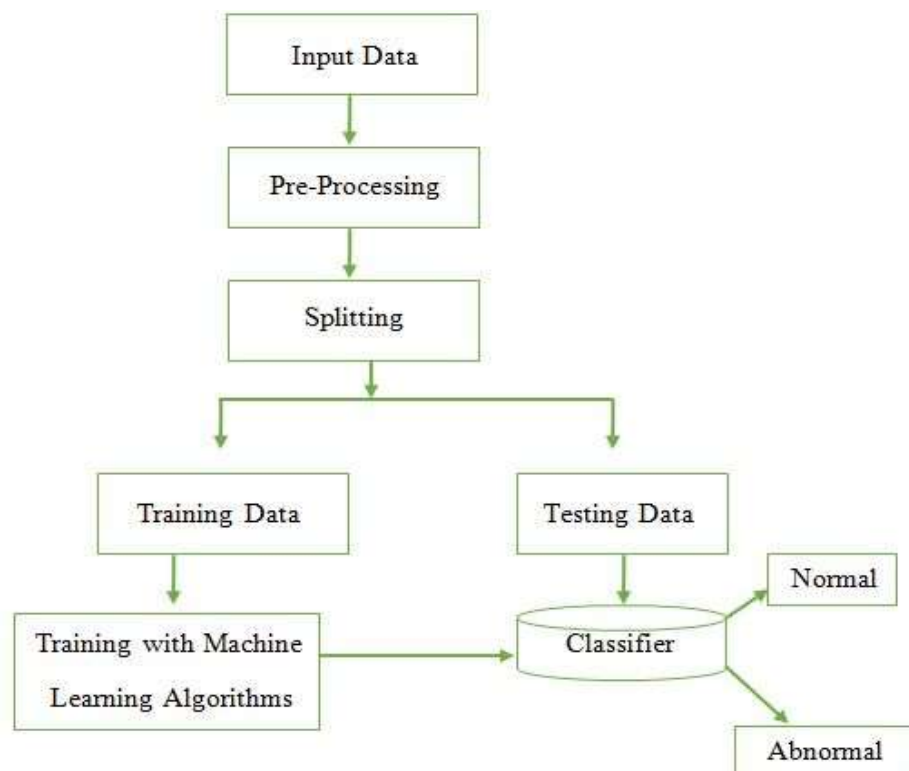
PROPOSED WORK

The most crucial task in the healthcare field is disease diagnosis. If a disease is diagnosed early, many lives can be saved. Machine learning classification techniques can significantly benefit the medical field by providing an accurate and quick diagnosis of diseases. Hence, save time for both doctors and patients. As heart disease is the number one killer in the world today, it becomes one of the most difficult diseases to diagnose. In this research work we examine the different machine learning algorithms for heart disease detection and compare the results using different performance measures, i.e. accuracy, precision, recall.

OBJECTIVES

1. Dataset collection which contains patient details.
2. Attributes selection process selects the useful attributes for the prediction of heart disease.
3. After identifying the available data resources, they are further selected, cleaned, made into the desired form.
4. Different Machine learning techniques as stated will be applied on pre-processed data to predict the accuracy of heart disease.
5. Accuracy measure compares the accuracy of different classifiers.

METHODOLOGY



CONCLUSION AND FUTURE WORK

Based on the above analysis, it can be concluded that there is considerable room for machine learning algorithms in predicting cardiovascular diseases or cardiac diseases. We have also studied various classification algorithms in the above paper that can be used to classify heart disease databases, and we have seen different techniques that can be used for classification and the accuracy obtained by them. An example would be-suppose the patient has diabetes that can cause heart disease in the future and then treat the patient with diabetes that can avoid heart disease in turn.

REFERENCES

- [1] BeantKaur, Williamjeet Singh proposed Review on Heart Disease Prediction System using Data Mining Techniques. International Journal on Recent and Innovation Trends in Computing and Communication,2014.
- [2] Sharma Purushottam, proposed review on “Heart Disease Prediction System Evaluation using C4.5 Rules and Partial Tree”, Springer, Computational Intelligence in Data Mining, vol.2, 2015.
- [3] Boshra Brahmi proposed review on “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, Journals of Multidisciplinary Engineering Science and Technology, vol.2, 2 February 2015
- [4] Nimai Chand Das Adhikari, Arpana Alka, and rajat Garg, “HPPS: Heart Problem Prediction System using Machine Learning”
- [5] <https://www.medicalnewstoday.com/articles/257484.php>.
- [6] Marjia Sultana proposed review on “Heart Disease Prediction using WEKA tool and 10-Fold cross-validation”, The Institute of Electrical and Electronics Engineers, March 2016.
- [7] K.Gomathi proposed review on “Multi Disease Prediction using Data Mining Techniques”, International Journal of System and Software Engineering, December 2016.
- [8] Dr.S.SeemaShedole proposed review on “Predictive analytics to prevent and control chronic disease”, <https://www.researchgate.net/publication/316530782>, January 2016.

- [9] Prajakta Ghadge proposed review on “Intelligent Heart Disease Prediction System using Big Data”, International Journal of Recent Research in Mathematics Computer Science and Information Technology, vol.2, October 2015 - March 2016.
- [10] Mr.P.Sai Chandrasekhar Reddy, Mr.PuneetPalagi, S.Jaya, “Heart Disease Prediction using ANN Algorithm in Data Mining”, International Journal of Computer Science and Mobile Computing, April 2017.
- [11] Zhao, J., & Wang, T. (2010). A general framework for medical data mining. *2010 International Conference on Future Information Technology and Management Engineering*, 2, 163-165.
- [12] Stephen F Weng machine learning improves cardio vascular risk prediction using routine clinical data (2017).
- [13] Himanshu Sharma Prediction of Heart Disease using Machine Learning Algorithms: A Survey (August 2017).
- [14] R. Sharmila proposed review on “A conceptual method to enhance the prediction of heart diseases using the data techniques”, International Journal of Computer Science and Engineering, May 2018.
- [15] Abhay Kishore, Department of Computer Engineering, Army Institute of Technology, Pune, Maharashtra Professor, Department of Computer Engineering, Army Institute of Technology, Pune, Maharashtra Heart Attack Prediction Using Deep Learning (2018).
- [16] M.Nikhil Kumar, Department of CSE, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India Prediction Heart Diseases using Data mining and machine learning algorithms and tools. (2018).
- [17] Mr.Chala Beyene proposed review on “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques”, International Journal of Pure and Applied Mathematics, 2018.