

A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach

M. Marimuthu
Assistant Professor
Coimbatore Institute of
Technology
Coimbatore

M. Abinaya
UG Scholar
Coimbatore Institute of
Technology
Coimbatore

K. S. Hariesh
UG Scholar
Coimbatore Institute of
Technology
Coimbatore

K. Madhankumar
UG Scholar
Coimbatore Institute of Technology
Coimbatore

V. Pavithra
UG Scholar
Coimbatore Institute of Technology
Coimbatore

ABSTRACT

Heart is the next major organ comparing to brain which has more priority in Human body. It pumps the blood and supplies to all organs of the whole body. Prediction of occurrences of heart diseases in medical field is significant work. Data analytics is useful for prediction from more information and it helps medical centre to predict of various disease. Huge amount of patient related data is maintained on monthly basis. The stored data can be useful for source of predicting the occurrence of future disease. Some of the data mining and machine learning techniques are used to predict the heart disease, such as Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K-Nearest Neighbour(KNN), Naive Bayes and Support Vector Machine (SVM). This paper provides an insight of the existing algorithm and it gives an overall summary of the existing work.

Keywords

Data mining, Heart disease, Machine learning, Medical centre.

1. INTRODUCTION

Heart disease is one of the prevalent disease that can lead to reduce the lifespan of human beings nowadays. Each year 17.5 million people are dying due to heart disease [1]. Life is dependent on component functioning of heart, because heart

is necessary part of our body. Heart disease is a disease that affects on the function of heart [2]. An estimate of a person's risk for coronary heart disease is important for many aspects of health promotion and clinical medicine. A risk prediction model may be obtained through multivariate regression analysis of a longitudinal study [3]. Due to digital technologies are rapidly growing, healthcare centres store huge amount of data in their database that is very complex and challenging to analysis. Data mining techniques and machine learning algorithms play vital roles in analysis of different data in medical centres. The techniques and algorithms can be directly used on a dataset for creating some models or to draw vital conclusions, and inferences from the dataset. Common attributes used for heart disease are Age, Sex, Fasting Blood Pressure, Chest Pain type, Resting ECG(test that measures the electrical activity of the heart), Number of major vessels colored by fluoroscopy, Threst Blood Pressure (high blood pressure), Serum Cholestrol (determine the risk for developing heart disease), Thalach (maximum heart rate achieved), ST depression (finding on an electrocardiogram, trace in the ST segment is abnormally low below the baseline), painloc (chest pain location (substernal=1, otherwise=0)), Fasting blood sugar, Exang (exercise included angina), smoke, Hypertension, Food habits, weight, height and obesity[4]. Table 1 summarizes the most common types of the heart disease as follows.

Table 1 Different types of heart disease [5]

Arrhythmia	The heart beat is improper whether it may irregular, too slow or too fast.
Cardiac arrest	An unexpected loss of heart function, consciousness and breathing occur suddenly.
Congestive heart failure	The heart does not pump blood as well as it should, it is the condition of chronic.
Congenital heart disease	The heart's abnormality which develops before birth.
Coronary artery disease	The heart's major blood vessels can damage or any disease occurs in the blood vessels.
High Blood Pressure	It has a condition that the force of the blood against the artery walls is too high.
Peripheral artery disease	The narrowed blood vessels which reduce flow of blood in the limbs, is the circulatory condition.
Stroke	Interruption of blood supply occur damage to the brain.

Figure 1 depicts the parts of human heart such as Left atrium, Right atrium, Right ventricle, Left ventricle, Aorta, pulmonary vein, Pulmonary valve, Pulmonary artery,

Tricuspid valve, Aortic valve, Mitral valve, Superior vena cava and Inferior vena cava.

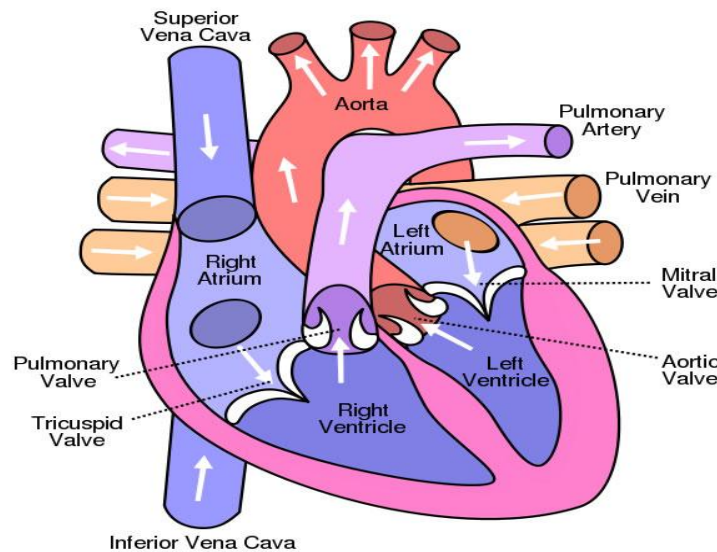


Figure 1 Human Heart [6]

This paper is organized as follows. Section 2 gives an overall literature review of the existing work. Section 3 provides a conclusion and future work.

2. LITERATURE REVIEW

There are numerous works has been done related to disease prediction systems using different data mining techniques and machine learning algorithms in medical centres.

K. Polaraju et al, [7] proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

Marjia et al, [8] developed heart disease prediction using KStar, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than KStar, Multilayer perception and J48 techniques using k-fold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory. Therefore, the accuracy's performance is improved more to give better decision to diagnosis disease.

S. Seema et al,[9] focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

Ashok Kumar Dwivedi et al, [10] recommended different algorithms like Naïve Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms.

MeghaShahi et al, [11] suggested Heart Disease Prediction System using Data Mining Techniques. WEKA software used for automatic diagnosis of disease and to give qualities of services in healthcare centres. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms.

Chala Beyene et al, [12] recommended Prediction and Analysis the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in short time. The proposed methodology is also critical in healthcare organisation with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex are some of the attributes are included to identify if the person has heart disease or not. Analyses of dataset are computed using WEKA software.

R. Sharmila et al, [13] proposed to use non- linear classification algorithm for heart disease prediction. It is proposed to use bigdata tools such as Hadoop Distributed File System (HDFS), Mapreduce along with SVM for prediction of heart disease with optimized attribute set. This work made an investigation on the use of different data mining techniques for predicting heart diseases. It suggests to use HDFS for storing large data in different nodes and executing the prediction algorithm using SVM in more than one node simultaneously using SVM. SVM is used in parallel fashion which yielded better computation time than sequential SVM.

Jayami Patel et al, [14] suggested heart disease prediction using data mining and machine learning algorithm. The goal of this study is to extract hidden patterns by applying data mining techniques. The best algorithm J48 based on UCI data has the highest accuracy rate compared to LMT.

Purushottam et al, [15] proposed an efficient heart disease prediction system using data mining. This system helps medical practitioner to make effective decision making based on the certain parameter. By testing and training phase a

certain parameter, it provides 86.3% accuracy in testing phase and 87.3% in training phase.

K.Gomathi et al, [16] suggested multi disease prediction using data mining techniques. Nowadays, data mining plays vital role in predicting multiple disease. By using data mining techniques the number of tests can be reduced. This paper mainly concentrates on predicting the heart disease, diabetes and breast cancer etc.,

P.Sai Chandrasekhar Reddy et al, [17] proposed Heart disease prediction using ANN algorithm in data mining. Due to increasing expenses of heart disease diagnosis disease, there was a need to develop new system which can predict heart disease. Prediction model is used to predict the condition of the patient after evaluation on the basis of various parameters like heart beat rate, blood pressure, cholesterol etc. The accuracy of the system is proved in java.

Ashwini shetty et al, [18] recommended to develop the prediction system which will diagnosis the heart disease from patient's medical dataset. 13 risk factors of input attributes have taken into account to build the system. After analysis of the data from the dataset, data cleaning and data integration was performed.

Jaymin Patel et al, [19] suggested data mining techniques and machine learning to predict heart disease. There are two objectives to predict the heart system. 1. This system not assume any knowledge in prior about the patient's records. 2. The system which chosen must be scalar to run against the large number of records. This system can be implemented using WEKA software. For testing, the classification tools and explorer mode of WEKA are used.

Boshra Brahmi et al, [20] developed different data mining techniques to evaluate the prediction and diagnosis of heart disease. The main objective is to evaluate the different classification techniques such as J48, Decision Tree, KNN, SMO and Naïve Bayes. After this, evaluating some performance in measures of accuracy, precision, sensitivity, specificity are evaluated and compared. J48 and decision tree gives the best technique for heart disease prediction.

Noura Ajam [21] recommended artificial neural network for heart disease diagnosis. Based on their ability, Feed forward Back propagation learning algorithms have used to test the model. By considering appropriate function, classification accuracy reached to 88% and 20 neurons in hidden layer. ANN shows result significantly for heart disease prediction.

Prajakta Ghadge et al, [22] suggested big data for heart attack prediction. The objective of this paper is to provide prototype using big data and data modelling techniques. It can be also

used to extract patterns and relationships from database which associated with heart disease. This system consists of two databases namely, original big dataset and another is updated one. A java-file system named HDFS used to provide a user with reliable. This system can assist the healthcare practitioners to make intelligent decisions. The automation in this system would be advantageous.

S.Prabhavathi et al, [23] proposed Decision tree based Neural Fuzzy System (DNFS) technique to analyse and predict of various heart disease. This paper reviews the research on heart disease diagnosis. DNFS stand for Decision tree based Neural Fuzzy System. This research is to create an intelligent and cost effective system, and also to improve the performance of the existing system. Specifically in this paper, data mining techniques are used to enhance heart disease prediction. The result of this research shows that the SVM and neural networks results highly positive manner to predict heart disease. Still the data mining techniques are not encouraging for heart disease prediction.

Sairabi H.Mujawar et al, [24] used k-means and naïve bayes to predict heart disease. This paper is to build the system using historical heart database that gives diagnosis. 13 attributes have considered for building the system. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used. 13 attributes with total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

Sharan Monica.L et al[25] proposed an analysis of cardiovascular disease. This paper proposed data mining techniques to predict the disease. It is intend to provide the survey of current techniques to extract information from dataset and it will useful for healthcare practitioners. The performance can be obtained based on the time taken to build the decision tree for the system. The primary objective is to predict the disease with less number of attributes.

Sharma Purushottam et al, [26] proposed c45 rules and partial tree technique to predict heart disease. This paper can discover set of rules to predict the risk levels of patients based on given parameter about their health. The performance can be calculated in measures of accuracy classification, error classification, rules generated and the results. Then comparison has done using C4.5 and partial tree. The result shows that there is potential prediction and more efficient. Table 2 describes the accuracy of the heart disease with different techniques are shown below.

Table 2 A comparative study of various algorithms in literature review.

YEAR	AUTHOR	PURPOSE	TECHNIQUES USED	ACCURACY
2015	Sharma Purushottam et al,[15]	Efficient Heart Disease Prediction System using Decision Tree.	Decision tree classifier	86.3% for testing phase. 87.3% for training phase.
2015	Boshra Brahmi et al, [20]	Prediction and Diagnosis of Heart Disease by Data Mining Techniques.	J48, Naïve Bayes, KNN, SMO	J48 gives better accuracy than other three techniques.
2015	Sairabi H. Mujawar et al, [24]	Prediction of Heart Disease using Modified K-means and by using	Modified k-means algorithm, naïve bayes algorithm.	Heart Disease detection=93%. Heart Disease

		Naïve Bayes.		undetected=89%.
2015	Noura Ajam et al, [21]	Heart Disease Diagnoses using Artificial Neural Network.	ANN	88%
2015	Sharma Purushottam et al, [26]	Heart Disease Prediction System Evaluation using C4.5 Rules and Partial Tree.	C4.5 rules and Naive Bayes algorithm	C4.5 gives better accuracy than Naive Bayes.
2016	Marjia et al, [8]	Prediction of Heart Disease using WEKA tool.	K Star	75%
			J48	86%
			SMO	89%
			Bayes Net	87%
			Multilayer Perception	86%
2016	S. Seema et al, [9]	Chronic Disease Prediction by mining the data.	Naïve Bayes	Highest accuracy achieved by SVM, in case of heart disease 95.56%
			Decision Tree	Highest accuracy of 73.588% achieved by Naïve Bayes in case of diabetes.
			Support Vector Machine	
2016	Ashok Kumar Dwivedi et al[10]	Evaluate the performance of different machine learning techniques for heart disease prediction.	Naïve Bayes	83%
			KNN	80%
			Logistic Regression	85%
			Classification Tree	77%
2016	K. Gomathi et al,[16]	Multi Disease Prediction using Data Mining Techniques.	Naïve Bayes	Heart Disease: 79% Diabetes: 77.6% Breast Cancer: 82.5%
			J48	Heart Disease: 77% Diabetes: 100% Breast Cancer: 75.5%
2016	Jayamin Patel et al, [19]	Heart Disease Prediction using Machine Learning and Data Mining Technique.	J48, Logistic model tree algorithm, Random forest algorithm	J48 gives 56.76% which is better than LMT algorithm of accuracy 55.75%.
2016	Ashwini Shetty A et al, [18]	Different Data Mining Approaches for Predicting Heart Disease.	WEKA tool, MATLAB. Neural Network	84%
			Hybrid Systems	89%
2016	Prajakta Ghadge et al, [22]	Intelligent Heart Disease Prediction System using	Hadoop, Mahout, Naïve bayes.	The automation of this system makes extremely

		Big Data.		advantageous.
2016	S. Prabhavathi et al, [23]	Analysis and Prediction of Various Heart Diseases using DNFS Techniques.	Decision tree, c4.5, SVM, naïve bayes.	Accuracy according to the types of heart disease. CVD Diagnosis= between 85% and 99%. CHD Diagnosis= between 82% and 92%.
2016	Sharan Monica. L et al,[25]	Analysis of CardioVascular Disease Prediction using Data Mining Techniques.	J48	91.4%
			Naïve Bayes	88.5%
			Simple CART	92.2%
2017	Jayami Patel et al,[14]	Heart disease Prediction using Machine Learning and Data mining Technique.	LMT, UCI	UCI gives better accuracy, compared to LMT.
2017	P. Sai Chandrasekhar Reddy et al, [17]	Heart disease prediction using ANN algorithm in data mining.	ANN	Accuracy proved in JAVA.
2018	Chala Bayen et al,[12]	Prediction and Analysis the occurrence of Heart Disease using data mining techniques.	J48, Naïve Bayes, Support Vector Machine.	It gives short time result which helps to give quality of services and reduce cost to individuals.
2018	R. Sharmila et al, [13]	A conceptual method to enhance the prediction of heart diseases using the data techniques.	SVM in parallel fashion	SVM provides better and efficient accuracy of 85% and 82.35%. SVM in parallel fashion gives better accuracy than sequential SVM.

3. CONCLUSION AND FUTURE WORK

By using different types of data mining and machine learning techniques to predict the occurrence of heart disease have summarized. Determine the prediction performance of each algorithm and apply the proposed system for the area it needed. Use more relevant feature selection methods to improve the accurate performance of algorithms. There are several treatment methods for patient, if they once diagnosed with the particular form of heart disease. Data mining can be of very knowledge form such suitable dataset.

In conclusion, as identified through the literature survey, believe only a marginal success is achieved in the creation of predictive model for heart disease patients and hence there is a need for combinational and more complex models to increase the accuracy of the predicting the early onset of heart disease. With the more amount of data being fed into the database the system will be very intelligent.

There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. Due to time limitation, the following research / work need to be performed for the future. Would like to make use of testing different discretization techniques, multiple classifier voting technique and different decision tree types

namely information gain and gain ratio. Willing to explore different rules such as association rule, logistic regression and clustering algorithms.

4. REFERENCES

- [1] Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017, pp.2137-2159.
- [2] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.
- [3] Guizhou Hu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Science of Cardiology, 10 May 2005.
- [4] T.Mythili, Dev Mukherji, Nikita Padaila and Abhiram Naidu, "A Heart Disease Prediction Model using SVM-Decision Trees- Logistic Regression (SDL)", International Journal of Computer Applications, vol. 68, 16 April 2013.

- [5] <https://www.medicalnewstoday.com/articles/257484.php>.
- [6] Nimai Chand Das Adhikari, Arpana Alka, and rajat Garg, “HPPS: Heart Problem Prediction System using Machine Learning”.
- [7] K. Polaraju, D. Durga Prasad, “Prediction of Heart Disease using Multiple Linear Regression Model”, *International Journal of Engineering Development and Research Development*, ISSN:2321-9939, 2017.
- [8] Marjia Sultana, Afrin Haider, “Heart Disease Prediction using WEKA tool and 10-Fold cross-validation”, *The Institute of Electrical and Electronics Engineers*, March 2017.
- [9] Dr.S.Seema Shedole, Kumari Deepika, “Predictive analytics to prevent and control chronic disease”, <https://www.researchgate.net/publication/316530782>, January 2016.
- [10] Ashok kumar Dwivedi, “Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation”, Springer, 17 September 2016.
- [11] Megha Shahi, R. Kaur Gurm, “Heart Disease Prediction System using Data Mining Techniques”, *Orient J. Computer Science Technology*, vol.6 2017, pp.457-466.
- [12] Mr. Chala Beyene, Prof. Pooja Kamat, “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques”, *International Journal of Pure and Applied Mathematics*, 2018.
- [13] R. Sharmila, S. Chellammal, “A conceptual method to enhance the prediction of heart diseases using the data techniques”, *International Journal of Computer Science and Engineering*, May 2018.
- [14] Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, “Heart disease Prediction using Machine Learning and Data mining Technique”, March 2017.
- [15] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, “Efficient Heart Disease Prediction System”, 2016, pp.962-969.
- [16] K.Gomathi, Dr.D.Shanmuga Priyaa, “Multi Disease Prediction using Data Mining Techniques”, *International Journal of System and Software Engineering*, December 2016, pp.12-14.
- [17] Mr.P.Sai Chandrasekhar Reddy, Mr.Puneet Palagi, S.Jaya, “Heart Disease Prediction using ANN Algorithm in Data Mining”, *International Journal of Computer Science and Mobile Computing*, April 2017, pp.168-172.
- [18] Ashwini Shetty A, Chandra Naik, “Different Data Mining Approaches for Predicting Heart Disease”, *International Journal of Innovative in Science Engineering and Technology*, Vol.5, May 2016, pp.277-281.
- [19] Jaymin Patel, Prof. Tejal Upadhay, Dr.Samir Patel, “Heart Disease Prediction using Machine Learning and Data Mining Technique”, *International Journal of Computer Science and Communication*, September 2015-March 2016, pp.129-137.
- [20] Boshra Brahmi, Mirsaeid Hosseini Shirvani, “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, *Journals of Multidisciplinary Engineering Science and Technology*, vol.2, 2 February 2015, pp.164-168.
- [21] Noura Ajam, “Heart Disease Diagnoses using Artificial Neural Network”, *The International Insitute of Science, Technology and Education*, vol.5, No.4, 2015, pp.7-11.
- [22] Prajakta Ghadge, Vrushali Girme, Kajal Kokane, Prajakta Deshmukh, “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, pp.73-77.
- [23] S.Prabhavathi, D.M.Chitra, “Analysis and Prediction of Various Heart Diseases using DNFS Techniques”, *International Journal of Innovations in Scientific and Engineering Research*, vol.2, 1, January 2016, pp.1-7.
- [24] Sairabi H.Mujawar, P.R.Devale, “Prediction of Heart Disease using Modified K-means and by using Naïve Bayes”, *International Journal of Innovative research in Computer and Communication Engineering*, vol.3, October 2015, pp.10265-10273.
- [25] Sharan Monica.L, Sathees Kumar.B, “Analysis of CardioVascular Disease Prediction using Data Mining Techniques”, *International Journal of Modern Computer Science*, vol.4, 1 February 2016, pp.55-58.
- [26] Sharma Purushottam, Dr. Kanak Saxena, Richa Sharma, “Heart Disease Prediction System Evaluation using C4.5 Rules and Partial Tree”, Springer, *Computational Intelligence in Data Mining*, vol.2, 2015, pp.285-294.