# A Critical Review of Various Techniques for Heart Disease Prediction

Apurva Joshi[1], Er. Jitendra Dangra[2] and Dr. M. K. Rawat [3]

Master of Engineering (C.S.E.), Lakshmi Narain College of Technology, Indore Madhya Pradesh, India[1]

Asst. Prof, Lakshmi Narain College of Technology, Indore Madhya Pradesh, India[2]

Prof, Lakshmi Narain College of Technology, Indore Madhya Pradesh, India[3]

apurvajoshi2012@gmail.com[1], jitendra.it@lnctindore.com[2], drmkrawat@lnctindore.com[3]

**Abstract:** *The data mining and their different applications have become more popular now a day. A number of large and small scale applications are developed with the help of data mining techniques i.e. predictors, regulators, weather forecasting systems and business intelligence. There are two kinds of model available for it namely supervised and unsupervised. The performance and accuracy of the supervised data mining techniques are higher as compared to unsupervised techniques therefore in sensitive applications the supervised techniques are used for prediction and classification. The supervised learning based data mining techniques for heart disease prediction are analyzed in this paper.*

**Keywords:** *Data Mining, Supervised learning, Neural Network, Genetic algorithm, Naïve Bayes, Heart disease.*

## 1. INTRODUCTION

The data mining [1] is a process of analysis of the data and extraction of the essential patterns from the data. These patterns are used with the different applications for making decision making and prediction related task. The decision making and prediction is performed on the basis of the learning algorithms. The data mining algorithms supports both kinds of learning supervised and unsupervised. In unsupervised learning only the data is used for performing the learning and in supervised technique the data and the class labels both are required to perform the accurate training. In supervised learning the accuracy is maintained by creating the feedbacks form the class labels and enhance the classification performance by reducing the error factors from the learning model.

The proposed work is intended to investigate these techniques in the application of the predictions. Data mining techniques have been widely used in medical industry for decision support systems for prediction and diagnosis of many diseases with correct accuracy. These techniques have been very effective in designing medical support systems due to their ability to discover hidden patterns and relationships among medical data. One of the most important field in which this system can be used is heart disease prediction because the rate of death due to heart disease is rapidly increasing all over the world. Almost every system that predicts heart diseases use clinical datasets having parameters and inputs from complex tests conducted in laboratory. There is no any system which predicts the chances of occurring heart disease based on various factors such as age, family history, diabetes, hypertension, low cholesterol, smoking alcohol intake, obesity, physical inactivity ,high pulse rate etc. Heart disease patients have many of these visible risk factors which are common and can be used very effectively for diagnosis. System based on such risk factors will not only help medical professionals but it would also give patients a warning of their chance of having heart disease even before they visit a hospital or go for costly medical checkups. This technique has

two most successful data mining tools, neural networks and genetic algorithms. The hybrid system implementation uses the global optimization benefit of genetic algorithm for initialization of neural network weights. By using this hybrid system, the learning is fast, more stable and accurate in comparison to back propagation. The system was implemented in Matlab and it may predict the risk of having heart disease with an accuracy of 89%.

## 2. LITERATURE SURVEY

The Healthcare industry is an information rich industry, but unfortunately not all the data are mined which is required for discovering hidden patterns & effective decision making thus the hidden information cannot be used efficiently. Advanced data mining techniques are used for knowledge discovery in database and for medical research, particularly in prediction of Heart disease. Chaitrali S. Dangare et al [2] has analyzed prediction systems for Heart disease analysis by using more number of input attributes. The system uses different attributes such as sex, blood pressure, cholesterol level, diabetes, pulse rate, etc like 13 attributes for the prediction of the likelihood of patient getting a Heart disease. Until now, 13 attributes have been used for heart disease prediction. In this research paper two more attributes have been added i.e. smoking and obesity. The techniques used for data mining classification, namely Decision Trees, Naïve Bayes, and Neural Networks have been analyzed on database of Heart disease. As per result by the comparison of the performance of these techniques, accuracy of Neural Networks, Decision Trees, and Naive Bayes are 100%, 99.62%, and 90.74% respectively. Thus, it may be concluded that out of these three classification models Neural Networks predicts Heart disease with highest accuracy.

The successful applications of data mining in various fields like e-commerce, retail and marketing has led to its application in various other sectors. Among these sectors a growing sector is healthcare. The healthcare environment is rich in information but having a lack of knowledge. There is a wealth of data available within the healthcare systems. But due to the lack of effective analysis tools to discover hidden relationships and trends in data the information is not efficiently utilized. JyotiSoni et al [3] intends to provide a survey of efficient techniques at present of knowledge discovery in databases using data mining that can be efficiently used in today's medical research particularly in prediction of heart disease. Various experiments have been performed to compare the performance of predictive data mining technique on the same datasets attributes and the outcomes of these experiments show that Decision Tree performs well and sometime Bayesian classification also performs with an accuracy as of decision tree but other predictive methods like Neural Networks, Classification based on clustering and KNN do not perform well as of decision tree. Another thing which can be concluded from the results is that the accuracy of the Decision Tree and Bayesian Classification can be improved after applying genetic algorithm to it so that the actual data size get reduced to get the optimal subset of attributes that is sufficient for heart disease prediction.

The main objective of Shadab Adam Pattekari and AsmaParveen [4] is to develop an Intelligent System using data mining modeling technique, namely, Naive Bayes. It is a user friendly wed based system. In this system an user answers the predefined questions. It retrieves the hidden data from the previously stored database and then compares the user answers with trained data sets to analyze the result. It can answer complex queries for the prediction of heart disease and thus helps healthcare practitioners to make intelligent decisions which cannot be performed by traditional decision support systems. Other benefit of this system is that by providing effective treatments at early stage, it also helps to reduce treatment costs.

The healthcare industry consists a huge amount of healthcare data which, are not "mined" to discover hidden information for effective decision making and thus cannot be efficiently used. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can be used to redress this situation. N. Aditya Sundar et al [5] describes a prototype by using data mining techniques, known as Naïve Bayes and WAC (weighted associative classifier).This system is able to answer complex 'what if' queries which cannot be answered by the traditional systems .Using medical profile of a patient such as sex, age, blood pressure and blood sugar level it can predict the likelihood of

patients to get a heart disease. It enables significant knowledge, like patterns, and relationships between medical factors related to heart disease, to be established. It can be used as a very effective training tool to train medical students, medical staff, and nurses to diagnose patients which have the symptoms for the occurrence of heart disease. It is a web based application and thus can be used efficiently in hospitals if they are having data ware house for their hospital. At present we are analyzing the performance of the two classification techniques of data mining by using various performance measures.

Data mining is the computer based process of extracting meaningful data from the set of large amount of data to use the hidden information effectively and efficiently. Data mining tools can be used to predict the future trends, allowing business to make effective, knowledge based and proactive decisions. Data mining tools can be used to answer business questions that the traditional database system takes much time to resolve. The huge amount of data is generated for heart disease analysis that is too complex and voluminous to be processed and extracted by traditional approaches. Data mining provides the technology and methodology to transform this large amounts of data into meaningful information for effective decision making. By using data mining techniques the time taken is very less for the prediction of the heart disease with more accuracy. In this paper R. Thanigaivel et al [6] survey various papers in which one or more algorithms of data mining are used for heart disease prediction. Using neural networks the result is nearly 100%. So it can be concluded that the prediction by using data mining algorithm gives optimized results. Applying data mining techniques to heart disease analysis it can provide as reliable performance as that is achieved in diagnosing heart disease by traditional approaches.

M.I. López et al [7] proposes a classification via clustering approach for the prediction of final result in a university course which is based on forum data. The objective is two main fields: to determine if student participation in the course forum can be a good predictor to analyze the final marks for the course and to examine that the proposed classification via clustering approach can obtain a similar accuracy to the traditional classification algorithms or not. Experiments were carried out using real dataset from university students of first year. Various clustering algorithms based on the proposed approach were compared with traditional classification algorithms to predict the result of student whether students pass or fail the course on the basis of their forum usage data. The result of the experiment show that the Expectation-Maximisation (EM) clustering algorithm yields similar result to those of best classification algorithms, especially when using only particular groups of datasets. Finally, the centroid of the EM clusters were describe to show the relationship between the two clusters and the two classes of students.

## 3. CONCLUSION

The data mining is helpful for analysis the data, when the manually analysis of the data is not feasible then the data mining techniques are applied for analysis. The data mining techniques are the computer based algorithms which identify the relationship among the data and extraction of the similar pattern data on which they are trained. This paper presented a critical review of various data mining based techniques for the classification and prediction of heart disease related data.

## REFERENCES

[1] Data Mining - Applications & Trends, http://www.tutorialspoint.com/data_mining/dm_applications_trends.htm.

[2] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction Systemusing Data Mining Classification Techniques", International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.

[3] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications (0975–8887) Volume 17– No.8, March 2011.

[4] Shadab Adam Pattekari and Asma Parveen, "Prediction System for Heart Disease Using Naive Bayes", International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624. Vol 3, Issue 3, 2012, pp 290-294.

[5] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra, "Performance Analysis of Classification Data Mining Techniques Over Heart Disease Data Base", International Journal of Engineering Science & Advanced Technology, Volume-2, Issue-3, 470 – 478.

[6] R. Thanigaivel, Dr. K. Ramesh Kumar, "Review on Heart Disease Prediction System using Data Mining Techniques",

Asian Journal of Computer Science and Technology (AJCST) Vol. 3. No.1 2015 pp 68-74.

[7] M. I. López, J. M Luna, C. Romero, S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums", Proceedings of the 5th International Conference on Educational Data Mining..