



A Review - Heart Disease Forecasting Pattern using Various Data Mining Techniques

¹Ramandeep Kaur, ²Er. Prabhsharn Kaur

¹M.tech Scholar CSE Dept, NWIET, Dhudike, India

²Assistant Professor CSE Dept, NWIET, Dhudike, India

deepdeol@hotmail.com, prabhsharanbrar@hotmail.com

Abstract - Data mining is the process of collecting large sets of data and then separating the meaning of data. Data mining provides the various techniques and methods for the transformation of data into useful information used for decision making in future. The data mining techniques namely Decision Tree, Naïve Byes, Neural Network, K-means clustering, association classification, Support vector machine(SVM) and MAFIA algorithm are analyzed on heart disease database. Heart disease defines a various healthcare conditions that are highly vast in nature which are related to heart and has many basic causes that affect the entire body. In this paper, we survey different papers in which one or more algorithms of data mining are used for prediction of heart disease. This paper discussed the overall review of heart disease prediction using various data mining techniques. These techniques used in heart disease take less time and make process fast for the prediction system to predict heart disease with good accuracy in order to improve health.

Keywords - Data mining, Heart Disease, Decision tree, K- means clustering, association classification, SVM, MAFIA

I. INTRODUCTION

In our busy schedule, most of the people work like a machine in order to live a deluxe and comfortable life in future and to earn more money, but during this type of situation people forget about their health and even don't take any proper rest. Because of this they affected from various type of diseases at a very early age, which cause our health as Diabetes, Cancer, Heart Disease and many more.[2] Now a day's heart disease is a major health problem and cause of death all over the world. Heart is a very valuable part of our body, and plays a very important role in our life. Our whole life depends on efficient working of heart. Heart disease is caused due to narrowing or blockage of coronary arteries. This is caused by deposition of fat on inner walls of arteries and also due to build up cholesterol. There are some of major heart disease factors which include Diabetes, high blood pressure, obesity, smoking, eating habits, alcohol that affects our whole body. [1]

Some of major symptoms of heart attack are:

- Fullness and pain in center of your chest.
- Pain or discomfort on one or both arms, the back, neck or stomach.
- Shortness of breath with or without chest discomfort.
- Nausea, Indigestion, Heartburn, or stomach pain.

- Sweating and Fatigue.
- Pressure in the upper back

Types of heart diseases:

Heart means ‘cardio’. Hence all heart diseases concern to category of cardiovascular diseases.

Some types are:

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

Nowadays, many hospitals keep their present data in electronic form through some hospital database management system. These systems generate large volume of data on daily basis. This data may be in form of free text, structured as in databases or in form of images.[4] This data may be used to extract meaning information which may be used for decision making. Data mining which is one of the KDD (Knowledge discovery in database) concentrates on finding meaningful patterns from large datasets. These patterns can be further analyzed and the result can be used for further valuable decision making and analysis. It has helped to confirm the best prediction technique in terms of its accuracy and error rate on the specific dataset.[4]

The World Health organization has predicted that 12 million deaths happen worldwide, every year due to heart diseases. Heart diseases can kill one person every 34 seconds in the United States [3]. Researchers have been applied various Data mining techniques like Association rule mining, Clustering, Classification. By Using some data mining techniques heart disease prediction can be made simple by using various characteristic to find out whether the person suffers from heart attack or not, and it also takes less time to for the prediction and improve the medical diagnosis of diseases with good accuracy and minimizes the occurrence of heart attack [2].

II. LITERATURE SURVEY

Numerous work has been done related to heart prediction system by using various data mining techniques and algorithms by many authors. The aim of all is to achieve better accuracy and to make the system more efficient so that it can predict the chances of heart attack.

Divya Kundra and Navpreet kaur[1] presents “review on prediction system for heart diagnosis using data mining techniques” .The analysis shows that using different techniques and taking different number of attributes we get different accuracies for predicting heart diseases.

Neha chauhan and Nisha Gautam[2] in 2013 present an overview of heart disease prediction using data mining techniques. These techniques take less time and make process fast for prediction system to predict heart disease with good accuracy to improve their health.

M.A.Nishara Banu and B.Gomathy[3] used C4.5 algorithm, MAFIA and K-Means clustering in the year 2014 using 13 attributes in dataset and achieving 89% accuracy.

Sujata Joshi and Mydhili[4] K-Nair uses three classification techniques namely decision tree, Naïve Bayes, and K-nearest neighbour and shows the performance among them. It has helped to determine the best prediction technique in terms of its accuracy and error rate on specifying dataset.

J.Soni[6] presents on overview of heart diseases prediction for medical diagnosis. This research provided survey of current techniques of knowledge of discovery in databases that are used in medical research in heart disease prediction by using data mining techniques.

Ms.Ishtake[8] developed a prediction system for heart diagnosis using decision tree, Neural Network and Naïve Bayes techniques using 15 attributes in year 2013.

Nidhi Bhatlaet[9] projected the study of different data mining techniques that can be employed in automated heart disease prediction systems. The analysis shows that Neural Network with 15 attributes has shown the accuracy

Shashikant Ghumbri, Chetan Patil and Ashok Ghatol[10]uses an intelligent system based Support Vector Machine along with a radial basis fuction network is presented for diagnosis and shows the result that SVM can be successfully used for diagnosing heart disease.

III. APPROACHES USED FOR DATA MINING

(a) K-means Clustering: Collecting a set of objects in such a way that objects in the same group is more similar to each other than those in other groups. This algorithm groups the sets of data points in space into a predefined number of cluster[11]. This technique aims to partition the set of observation points in to K-clusters. In this information is clusters the information in to K-Groups where K is considered as input parameter and assigns each

information's to clusters based upon the observation's proximity to the mean of cluster. The cluster's mean is then more computed and the process will continue again[3].

Steps involved in K-means algorithm are as follows:

- Select the number of clusters.
- Randomly create k clusters and find center of cluster.
- Calculate the distance between each data point and cluster center.
- Assign the data point to the nearest cluster center.
- Recalculate the new cluster center.
- Repeat the above steps until the termination condition is met.

This technique is mainly used in various medical applications such as heart disease prediction. This technique allows running on large databases. This is the fastest algorithm among all.

(b) Decision Trees: This technique is more suitable for classification problem. Classification is an unsupervised learning used to predict the class of objects whose class label is unknown [3]. In this technique mainly two steps are involved making a tree and implement that tree to the dataset. There are various decision tree algorithms such as C4.5, ID3, C5.0, and CART [5].

C4.5 algorithm is an extension of the basic ID3 algorithm. This approach gives maximum accuracy on training data. Decision Tree Structure provides an "if-then" rule that makes the result easy for explanation or understand. This is generally used in various operation researches particularly in decision making and to identify a technique to achieve a goal [2].

(c) Support Vector Machine: Support vector machine is a classification and regression method used to maximize the predictive accuracy without over fitting the data to be trained. This method has the ability to resolve linear and non linear classification problems. This technique is mostly used for larger number of datasets, with a large number of predictor fields [12].

This technique is based on mathematical functions. This performs well on data sets having maximum number of attributes. SVM map the training data into Kernel space. Mainly Kernel spaces used are- linear(uses dot product), Quadratic, polynomial Radial Basis function kernel, Multilayer perception kernel etc There are also multiple methods to implement SVM such as quadratic programming sequential minimal optimization and least squares[13]. A disadvantage is that the algorithm is sensitive to the choice of variable settings, making it harder to use and time consuming to identify the best.

(d) Naive Bayes: This is a classifier approach which is used when the dimensionality of the input is high. This is a basic technique for many machine learning methods and used to create models. This technique provides a way of exploring and understanding the data. This method is based on Bayes theorem.

$$P(Y/X) = P(X/Y)/P(X)$$

This thus calculates the probability of Y given X, when Y represents the dependence event and X represents the prior event [2]. This algorithm is based on the assumption that an attribute value on a given class is independent of the values of other attributes [5].

IV. DATA MINING APPLICATIONS

Data mining is widely used in diverse areas. There are a number of data mining system available today. Some of these are:

(i) Medical Industry: Data /mining is very useful in healthcare industry in the diagnosis of heart disease, lung cancer, breast cancer and diabetes. The heart medical prediction system predicts the chances of heart attack to a person with better accuracy. Data mining can be used to predict the volume of patients in every category. The large amounts of data generated by health care industries are too complex so data mining provides the techniques to transform these sets of data so that it can be used for decision making.

(ii) Retail Industry: Data mining gathers huge amount of data from on sales, transportation, consumption and services which leads to a great application in retail industry. The quantity of data gathered will continue to expand for more popularity. Data mining leads to improve the service of customer and also customer satisfaction.

(iii) Intrusion Detection: Intrusion detection refers to some type of action that threatens the accuracy and integrity of resources. Security is one of the major issues which we have to maintain. As internet is used on a large scale and availability of tools and tricks for intruding and attacking network is prompted intrusion detection to become a critical component administration [1].

(iv) Education: Data mining can also be useful for education systems concerns with developing methods that discover knowledge from data originating from educational environments. This can be useful to evaluate the

performance of a student by analysis the total record of the students and helpful for institution to take accurate decisions. Learning pattern of the students can be collected and used to develop efficient methods to teach them.

(v) **Fraud detection:** Previous methods used for fraud detection are time consuming and complex. Data mining aids in providing meaningful patterns and converts data into useful information which can be used for decision making in future. A perfect fraud detection system should protect information of all the users. A method is used for the collection of all sample records. These records are classified as fraudulent or non-fraudulent. All this data is used to make a model and the algorithm is made to identify whether the given record is fraudulent or not.

V. CONCLUSION

A novel heart disease prediction system aims to predict the possibility of heart disease risk of patients. Analysis predicted by different researcher’s shows that using different data mining techniques having different number of attributes, we get different accuracies for predicting heart disease. This paper presents a survey for various data mining approaches such as decision making, K-means clustering, Naïve-Byes, Support vector machine(SVM) for heart disease prediction and compare the accuracies of different algorithms for heart disease prediction system. Table 1 represents the concluding view of various data mining techniques with multiple combinations.

Table1: Summarized view of various data mining techniques

Approaches used	No of attributes	Accuracy
K-means, MAFLA, C4.5	13	89%
KNN, Neural Network, Naïve Byes, Decision Tree	13	99.2%
Decision Tree, Naïve Byes, Neural Network	13	89.2%
Neural Network, SVM	13	85%

As various techniques are used for this purpose, but there is always a need for better approach to increase the accuracy of the system and to get better results.

REFERENCES

- [1] Divya Kundra and Navpreet Kaur, "Review on Prediction System for Heart Diagnosis using data mining techniques", International Journal of Latest Research in Engineering and Technology, volume 1, October 2015.
- [2] Neha Chauhan and Nisha Gautam "An Overview of heart disease prediction using data mining techniques",
- [3] M.A.Nishara Banu and B. Gomathy "Disease Forecasting System using Data Mining Systems", International Conference on Intelligent Computing Systems, 2014
- [4] Sujata Joshi and Mydhili K.Nair, "Prediction of Heart Disease Using Classification Based Data Mining Techniques", Springer India 2015, volume 2.
- [5] K.Sudhakar and Dr. M.Manimeklai "Study of Heart Disease prediction using data mining", International Journal of Advanced Research and Software Engineering, volume 4, January 2014.
- [6] Jyoti Soni and Suni Soni, "Predictive Data Mining for medical Diagnosis: An Overview of Heart Diseases prediction", International Journal of Computer Applications, volume 17, March 2011
- [7] Beant Kaur, Williamjeet Singh, "Review on Heart Disease Prediction System using Data mining Techniques" International Journal on Recent and Innovation trends in computing and Communication, vol2, No.10, 2014
- [8] Ms. Ishtake S.H. Prof. Snap S.A "Intelligent heart disease prediction system using data mining techniques", International J. of Health Care & Biomedical Research 2013.
- [9] Nidhi Bhatlet and Kiran Jyoti, "An Analysis of Heart disease prediction system using different data mining techniques", International Journal of Engineering Research and Technology, ISSN, volume 1, October-2012
- [10] Shashikant Ghumbri, Chetan Patil and Ashok Ghatol "Heart Disease diagnosis using Support Vector Machine", International Conference on Computer Science and Information Technology (ICCSIT'2011) Pattaya 2011
- [11] G.Komarasamy and Amitabh Wahi "Improving the cluster performance by combining PSO and K-Means algorithm", ICTACT Journal On Soft Computing, April 2011, Volume 01, Issue 04.

- [12] Brijendra Gupta and RB Mishra “ Neuro Psychiatric Disease Prediction using Support Vector Machine”, International Journal of Computational Bioinformatics Vol 2, No. 6 (2013).
- [13] Mythili T., Dev Mukherji, Nikita Padalia and Abhiram, “A Heart Disease Prediction Model using SVM-Decision Trees-Logistic regression(SDL)”, International Journal of computer applications(0975-8887) Volume 68-No.16(April 2013)
- [14] K.Srinivas, G.Raghavendra Rao and A. Govardhan “Prediction of Heart Morbidity Using data mining Techniques”, International Journal of Data Mining and Knowledge Management Process, vol1, No.3,2011.