ISSN:0975-9646

A Review on Prediction of Heart Diseases by Comparing Risk Factors in Data Mining

K.Manimekalai,

Head & Assistant Professor ,Department of Computer Applications,

Sri GVG Visalakshi college For Women, Udumalpet. TamilNadu, India

***Abstract* — Heart diseases are the number one cause of death. To protect the life of a patient from heart diseases there have to be quick and efficient detection technique to be followed. In heart disease prediction system, the input attributes play a major role for efficient prediction. Generally, many attributes were involved in predicting the heart disease. The objective of this paper is to study the risk factors in heart disease.**

***Keywords*— Data Mining, Heart Disease, Decision Tree, Naïve Bayes and Neural Network.**

I. INTRODUCTION

In modern society, millions of people suffer from heart disease annually. Heart disease is a major cause of morbidity and mortality in modern society. Medical diagnosis is extremely important but complicated task that should be performed accurately and efficiently. Although significant progress has been made in the diagnosis and treatment of heart disease, further investigation is still needed. The availability of huge amounts of medical data leads to the need for powerful data analysis tools to extract useful knowledge. There is a huge data available within the healthcare systems. According to the World Health Organization, 12 million deaths are caused by heart diseases in the world annually, 50 percent of which can be prevented by controlling risk factors. Heart diseases are expected to be the main reason for 35 to 60 percent of total deaths expected worldwide by 2025.

Adults who use proton pump inhibitors are between 16 and 21 percent more likely to experience a heart attack. A data-mining study has found an association between the use of proton-pump inhibitors, which account for 100 million prescriptions per year in the United States alone, and the likelihood of incurring a heart attack down the road.[7]

Proton pump inhibitors(PPIs) have been associated with adverse clinical outcomes amongst clopidogrel users after an acute coronary syndrome. Recent pre-clinical results suggest that this risk might extend to subjects without any prior history of cardiovascular disease. [7]

Data mining is the process of extracting hidden knowledge from data. It can reveal the patterns and relationships among large amount of data in a single or several datasets. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. In other words Data mining is one of the steps of knowledge discovery for extracting implicit patterns from vast, incomplete and noisy data. It is a field with the confluences of various disciplines that has brought statistical analysis, machine learning techniques, artificial intelligence and database management systems together to address the issues.

II. HEART DISEASE

The heart is important organ of human body part. It is nothing more than a pump, which pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops working altogether, death occurs within minutes. Life is completely dependent on efficient working of the heart. The term Heart disease refers to disease of heart & blood vessel system within it.

A number of factors have been shown that increases the risk of Heart disease :

- Family history
- Smoking
- Poor diet
- High blood pressure
- High blood cholesterol
- Obesity
- Physical inactivity
- Hyper tension

Factors like these are used to analyze the Heart disease. In many cases, diagnosis is generally based on patient's current test results & doctor's experience. Thus the diagnosis is a complex task that requires much experience & high skill.

II. LITERATURE SURVEY

Numerous studies have been done that have focus on diagnosis of heart disease. They have used different attributes and applied different data mining techniques for diagnosis & achieved different probabilities for different methods.

Swathi et al [1] proposed data mining tech niques that are used are Naive Bayes, Decision Tree and neural network. In that 12 attributes were used. The Naïve Bayes classifiers have the minimum error rate in comparison to all other classifiers. Bayesian classifiers have exhibited high accuracy and speed when applied to large databases. Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Decision trees can handle both categorical and numerical data. A neural network usually involves a large number of processors operating in parallel, each with its own small sphere of knowledge and access to data in its local memory. Typically, a neural network is initially "trained" or fed large

amounts of data and rules about data relationships. By applying these algorithms to the newer set of attributes, a more efficient system is developed.

Prachi Paliwal, Mahesh Malviya [2] introduced a new method based on the fitness value of the attribute to predict the heart disease problem. From the experiment it clear that proposed method is more accurately classify the recodes as compared to pervious method. Proposed method considers all attribute given to heart attack condition. It is also simple to understand and calculation is also easy. In that have only ten attributes which are mainly responsible for heart attack, in future more then ten attributes will increase which are also responsible for heart attack.

Boshra Bahrami et al [3] evaluated different classification techniques in heart disease diagnosis. Classifiers like J48 Decision Tree, K Nearest Neighbors (KNN), Naive Bayes (NB), and SMO are used to classify dataset. After classification, some performance evaluation measures like accuracy, precision, sensitivity, specificity, F-measure and area under ROC curve are evaluated and compared. The comparison results show that J48 Decision tree is the best classifier for heart disease diagnosis on the existing dataset. In that only 8 attributes were used.

Aqueel Ahmed et al [4] examined the classification techniques in data mining and show the performance of classification among them. In these classification accuracy among these data mining techniques has discussed. The result shows the difference in error rates. However there are relatively differences in different techniques. Decision tree and SVM perform classification more accurately than the other methods. In that paper 11 attributes were used. That may be used as reliable indicators to predict presence of heart disease. To increase the accuracy for the heart disease patient by increasing the various parameters suggested from the doctors by using different data mining techniques.

Vikas Chaurasia et al [4] employed bagging algorithm because it produce human readable classification rules which are easy to interpret. Researchers have been investigating applying different data mining techniques to help health care professionals in the diagnosis of heart disease. Bagging algorithm is one of the successful data mining techniques used in the diagnosis of heart disease patients. This paper investigates experiments are conducted to find the best classifier for predicting the diagnosis of heart disease patients. 11 attributes were used. This paper systematically investigates applying different methods of classifier technique in the diagnosis of heart disease patients. The results show that bagging algorithm accuracy of 85.03% and the total time taken to build the model is at .05 seconds in the diagnosis of heart disease patients.

Aqueel Ahmed, Shaikh Abdul Hannan [5] examined the classification techniques in data mining and shows the performance of classification among them. In these classification accuracy among these data mining techniques has discussed. The result shows the difference in error rates. In future, try to increase the accuracy for the heart disease patient by increasing the various parameters suggested from the doctors by using different data mining techniques.

Nidhi Bhatla et al [6] provided a study of different data mining techniques that can be employed in automated heart disease prediction systems. Various techniques and data mining classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that Neural Network with 15 attributes has shown the highest accuracy i.e. 100% so far. On the other hand, Decision Tree has also performed well with 99.62% accuracy by using 15 attributes. Moreover, in combination with Genetic Algorithm and 6 attributes, Decision Tree has shown 99.2% efficiency.

Jenn-LongLiu et al [8] presented two kinds of evolutionary data mining (EvoDM) algorithms, termed GA-KM and MPSO-KM, to cluster the dataset of cardiac disease and predict the accuracy of diagnostics. 13 attributes were used. The experiments indicated that the clustering accuracy of cardiac disease dataset is significantly improved by using GA-KM and MPSO-KM when compared to that of using K-means only.

III. PROPOSED PREDICTION SYSTEM

Today, most of the hospitals manage healthcare data using healthcare information system; the system contains large amount of data, used to extract hidden information for making intelligent medical diagnosis. Most of the researchers use the same attributes. In the modern society many persons affecting heart diseases. Day by day the risk factors are increased. To get more appropriate results, attributes should be increased. Table 1 shows the different input parameters.

Table 1: Dataset Description

S.No	Attribute	Type
1	Age	Real
2	Sex	Binary
3	Chest pain type	Nominal 4 values 1:typical type angina, 2:typical type angina, 3: non-angina pain; 4: asymptomatic
4	Resting Blood Pressure	Real
5	Serum cholesterol in mg/dl	Real
6	Fasting blood sugar >120 mg/dl	Binary
7	Diabetics	Real
8	Eating Habitsess History	Real
9	Stress	Binary
10	Old peak = ST depression induced by exercise relative to rest	Real
11	Family History	Real
12	thal: 3 = normal; 6 = fixed defect; 7 = reversible defect	Nominal
13	Slope	Ordered
14	Hypertension	Binary
15	Resting electro cardio graphic results(values 0,1,2)	Nominal

Some of Risk factors also responsible for Heart disease are obesity, smoking, Slope, Family history, Eating Habits, Stress, Diabetes, Sedenatary life cycle, type of job, high levels of uric acid, triglycerids, lipoprotein, fibrinogen, proton pump inhibitators.

Table 2 shows the number of attributes and Data Mining Techniques used in the existing systems.

Table 2 : Comparision of Attributes

Year	Author	Common Attributes used	Techniques used
2015	Swathi et al[1]	1,2,3,5,7,11	Naive Bayes, Decision Tree, Neural Network
2015	Prachi Paliwal et al [2]	1,2,4,5,8,9	Baysein classifier
2015	Boshra Bahrami et al[3]	1,3,4,6,15	J48, KNN, NB, SMO
2013	Vikas Chaurasia et al[4]	1,2,3,5,6,13	Naïve Bayes, J48, Bagging
2012	Aqueel Ahmed et al[5]	1,2,3,4,5,6	Decision tree, SVM
2012	Nidhi Bhatla et al [6]	1,3,4,15	Naïve Bayes, Decision Tree, KNN

Swathi et al used some common attributes like age, sex , Chest pain type, Serum cholesterol in mg/dl, Diabetics, Family History. The techniques used were Naive Bayes, Decision Tree,Neural Network.

Prachi Paliwal et al used common attributes like age, sex, Resting Blood Pressure, Serum cholesterol, eating habits and stress. Baysein classifier techniques was used.

Boshra Bahrami et al used the attributes age, Chest pain type, Resting Blood Pressure, Fasting blood sugar, Resting electro cardiographic results. The techniques J48, KNN, NB, SMO were utilised.

Vikas Chaurasia et al used the following attributes like age, sex, Chest pain type, Serum cholesterol, Fasting blood sugar, Slope. And also the techniques Naïve Bayes, J48, Bagging were utilized.

Aqueel Ahmed et al used age, sex, chest pain type, Blood Pressure, Cholestrol and Fasting blood sugar. The techniques Decision tree, SVM were used.

Nidhi Bhatla et al used the attributes like age, Chest pain type, Blood Pressure, Resting ECG. And the techniques
Naïve Bayes, Decision Tree, KNN were used.

Naïve Bayes, Decision Tree and KNN are the common techniques used to predict the heart disease.

IV RESULTS

The number of input attributes were varied. To get the accurate result, and for easy diagnosis and prediction of heart disease, the number of attributes should be standard.

Each author used various number of attributes. Swathi et al[1] used 12 attributes. Prachi Paliwal et al[2] used 12 attributes. Boshra Bahrami et al [3] used 8 attributes. Vikas Chaurasia et al[4] used 11 attributes. Aqueel Ahmed et al[5] used 11 attributes. Nidhi Bhatla et al[6] used 6 attributes. The number of input parameters are varied by authors.

Table 3 : Total number of Attributes used

Author	Number of Attributes used
Swathi et al[1]	12
Prachi Paliwal et al[2]	12
Boshra Bahrami et al [3]	8
Vikas Chaurasia et al[4]	11
Aqueel Ahmed et al[5]	11
Nidhi Bhatla et al[6]	6

V CONCLUSION

The objective of this work is to provide a study of different input attributes that can be used to predict the heart disease. Some attributes are mainly responsible for heart disease. In order to predict the disease based upon patient's raw symptom description, different input parameters can be collected. To help the physicians and healthcare professionals in the prediction of heart diseases, the number of attributes should standard. The system helps the medical practitioners to minimize the time of prediction and improve their practice as well as protect the life of the patient.

REFERENCES

[1] Swathi P, Yogish H.K, Sreeraj R, "Predictive Data Mining Procedures for the Prediction of Coronory Artery disease" , IJETAE Volume 5, Issue 2 February 2015.

[2] Prachi Paliwal, Mahesh Malviya, "An Efficient Method for Predicting Heart Disease Problem Using Fitness Value", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1290-1293.

[3] Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST), ISSN: 3159-0040, Vol. 2 Issue 2, February - 2015

[4] Vikas Chaurasia, Saurabh Pal, "Data Mining Approach to Detect Heart Diesease", International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol. 2, No. 4, 2013, Page: 56-66, ISSN: 2296-1739

[5] Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases: An Overview", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-1, Issue-4, September 2012

[6] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October - 2012 ISSN: 2278-0181

[7] Nigam H. Shah , Paea LePendu , Anna Bauer-Mehren, Yohannes T. Ghebremariam, Srinivasan V. Iyer, Jake Marcus, Kevin T. Nead, John P. Cooke, Nicholas J. Leeper, " Proton Pump Inhibitor Usage and the Risk of Myocardial Infarction in the General Population", PLOS ONE, June 2015.

[8] Jenn-LongLiu, Yu-Tzu, Chih-Lung Hung, "Development of evolutionary data mining algorithm and their applications to cardiac disease diagnosis", World congress on computational intelligence,2012.

[9] Yanwei, X.; Wang, J.; Zhao, Z.; Gao, Y., "Combination data mining models with new medical data to predict outcome of coronary heart disease". Proceedings International Conference on Convergence Information Technology 2007, pp. 868 – 872.

[10] http:// http://medicine.stanford.edu/.

[11] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", (IJERT) Vol. 1 Issue 8, October - 2012 ISSN: 2278-0181.