A Survey On Data Mining Techniques To Find Out Type Of Heart Attack

Prof. Gondkar Mayura D.[1], Prof. Pawar Suvarna E.[2]

(Department Of Information Technology, A.V.C.O.E, Sangamner , India)
(Department Of Information Technology, A.V.C.O.E, Sangamner, India)

 Abstract: *Heart disease is a major cause of morbidity and mortality in present society. Medicinal identification is extremely important but complicated task that should be performed precisely and proficiently. Although substantial advancement has been made in the diagnosis and treatment of heart disease, additional investigation is still needed. The obtainability of massive expanses of medical data leads to the need for powerful data analysis tools to extract useful information. There is a massive data available within the healthcare organisations. However, there is a task of effective analysis tools to discover trends in data , hidden relationships , Knowledge discovery and data mining have found numerous application in business and scientific domain. Investigators have long been apprehensive with applying statistical and data mining tools to improve data analysis on very big data sets. Any of the disease diagnosis is one of the applications where data mining tools are proving successful results. This paper projected to find out the heart attack type through data mining.*
Keywords: *Data Mining, Heart Disease*

I. INTRODUCTION

1. 1 Overview Of Data Mining

The Knowledge discovery in databases is a well-defined process which consists of numerous distinct steps. Data mining is the essential step, which results in the discovery of hidden but useful knowledge from substantial stores. A formal definition of Knowledge discovery in databases is given as follows: ―Data mining is the nontrivial extraction of implicit previously unknown and potentially useful information about data □ [1 l]. Data mining technology provides a user-oriented approach to novel and hidden forms in the information. The discovered data can be used by the healthcare administrators to improve the superiority of service. The discovered knowledge can also be used by the medical practitioners to reduce the number of adverse drug effect, to suggest less expensive therapeutically equivalent alternatives. Anticipating patient's future performance on the given history is one of the important advice less expensive medicinally equivalent alternatives. Anticipating patient's future performance on the given history is one of the important applications of data mining techniques that can be used in health care management.

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at reasonable costs. Quality provision implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions may lead to disastrous consequences which may be unacceptable. Hospitals should also minimize the cost of clinical tests. And these can be achieved by the results of employing appropriate computer-based information and/or decision support systems. Health care data is very massive. It includes patient centric data, resource management data and transmuted data. Health care organizations must have ability to examine data. Treatment records for millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care. The availability of combined information which comprises huge patient repositories, in which there is a shift in the awareness of clinicians, patients and spenders from qualitative visualization of the medical data by demanding a more quantitative assessment of information with the supporting of all clinical and imaging data. For example it might now be possible for the physicians to compare diagnostic information of various patients with identical conditions. Likewise, physicians can also confirm their findings with the conformity of other physicians dealing with an identical case from all over the world [12]. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The computerization of the same would be highly valuable.

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This repetition leads to undesirable preconceptions, mistakes and undeserved medical costs which affects the quality of service provided to patients. The Author name Wu, et al has proposed that integration of clinical decision support with computer based patient records could reduce the medical errors, which could automatically enhanced the patient safety, and will decrease unwanted practice variation, and improve patient outcome [13]. This suggestion is promising as data modeling and analysis tools,

namely data mining, has the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.
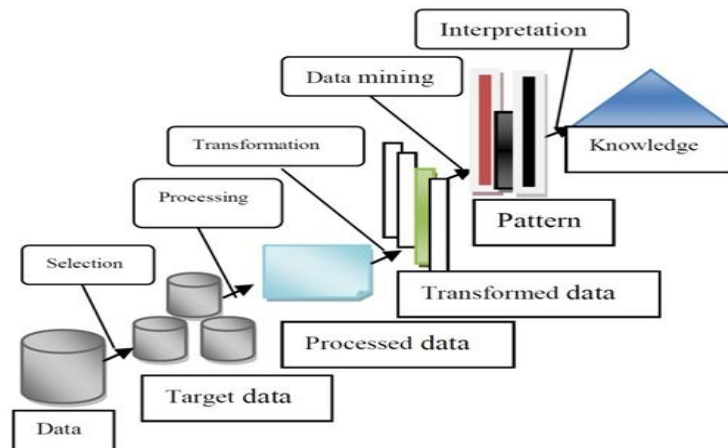

Fig 1.1: KDD Process

1.2 Cardiovascular Diseases

Cardiovascular disease is caused by the disorders of the heart and blood vessels, which leads to a disease called as Coronary Heart Disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure which leads to hypertension, peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure. The major causes of cardiovascular diseases are use of tobacco, physical inactivity, unhealthy diet and intake of alcohol. These are the three main causes which leads to heart diseases (1) chest pain (2) stroke and (3) heart attack.[15].To avoid and identification of these diseases different techniques of data mining is used through this easily find out heart related diseases specially Heart Attack using real patient data and this is the aim of the paper. Heart disease is the leading cause of death all over the world in the past ten years. Numerous researchers are using statistical and data mining tools to help health care professionals in the diagnosis of heart disease. [10]

1.3 Risk factors for heart disease
Following are some of the risk factors for heart disease:

1) Smoking: Smokers risk a heart attack twice as much as non smokers.

2) Cholesterol: A diet low in cholesterol and saturated Tran's fat will help lower cholesterol levels and reduce the risk of heart disease.

3) Blood pressure: High Blood Pressure leads to heart Attack

4) Diabetes: Diabetes if not controlled can lead to significant heart damage including heart attack and death

5) Sedentary life style: Simple relaxation time activities like gardening and walking can lower our risk of heart disease.

6) Eating Habits: Healthy diet, intake of low salt in diet, saturated fat in body, Tran's fat, cholesterol and refined sugars will lower our chances of getting heart disease. Stress: Poorly controlled stress an danger can lead to heart attacks and strokes

II.   RELATED WORK

Few research works has been carried out for diagnosis of various diseases using data mining. Our approach is to apply feature subset selection and artificial neural networks for prediction of heart disease.M.A.Jabbaret.al proposed a new algorithm combining associative classification and feature subset selection for heart disease prediction [4]. They applied symmetrical uncertainty of attributes and genetic algorithm to remove redundant attributes. Enhanced prediction of heart disease using genetic algorithm and feature subset selection was proposed by Anbarasiet.al [5]. Heart disease prediction using associative classification was proposed by M.A. Jabbar et.al [6]. Matrix based association rule mining for heart disease prediction was proposed by M.A. Jabbaret.al [7]. Association rule mining and genetic algorithm based heart disease prediction was proposed in [8]. Cluster based association rule mining for disease prediction was proposed in [14]. Sellappan Palaniappan et al proposed intelligent heart disease prediction system using naïve Bayes, decision tree and neural network in [1]. Graph based approach for heart disease prediction for Andhra Pradesh population was proposed by M.A.Jabbar et.al [8]. They combined maximum clique concept in graph with weighted association rule mining for disease prediction. Feature subset selection using FCBF in type II Diabetes patient's data was proposed by Sarojinibala Krishnan et.al. [9]. Heart disease prediction using

associative classification and genetic algorithm was proposed by M.A. Jabbaret.al [4].in their paper they used Z-Statistics a measure to filter out the rules generated by the system.

III.   DATASET

As per review of existing systems thirteen attributes are used for heart attack prediction. As the dataset consists of large number of attributes, the processing time and efforts required are more in enough. So there is need to reduce the number of attributes considered during overall processing to achieve same kind of results. Mostly standard data set is used to carry out the medical diagnosis, but we can also have better alternative to use analytical dataset instead of standard dataset. The resultant data set thus obtained may be used for the classification model to calculate the efficiency of the modified system.

Table 1. Description of 13 Input Attributes

Sr.No	Attribute	Description	Values
01	AGE	Age in years	Continuous
02	SEX	Male or Female	0=Male 1=Female
03	RESTECG	Resting electrographic results	Value 0 = Normal Value 1 = Having ST-T wave abnormal Value 2 = Left ventricular hypertrophy
04	RBP	Resting blood pressure	Continuous value in mm hg
05	SLOPE	Slope of the peak exercise ST segment	2 = Unsloping 3 = Flat 4 = Downsloping
07	FBS	Fasting blood sugar	$1 \geq 120$ mg/dl $0 \leq 120$ mg/dl
08	THALACH	Maximum heart rate achieved	Continuous value
09	CHOL	Serum cholesterol	Continuous value in mm/dl
10	EXANG	Exercise induced angina	0 = No 1 = Yes
11	OLDPEAK	ST depression induced by exercise relative to rest	Continuous value
12	CA	Number of major vessels colored by floursopy	0-3 value
	CP	Chest pain type	1 = Typical type 1 2 =Typical type angina 3 = Non-angina pain 4 = Asymptomatic
13	THAL	Defect type	2= Normal 6 = Fixed 7 =Rreversible defect

IV.   DATA MINING ALGORITHMS AND TECHNIQUES

Various algorithms and techniques like Classification, Clustering, Artificial Intelligence, Regression, Neural Networks, Association Rules, Genetic Algorithm, Decision Trees, Nearest Neighbour method etc., are used for knowledge discovery from databases

4.1. Classification

Classification is the most commonly applied data mining technique, which needs a set of pre-classified examples to develop a model that can classify the population of records at huge. Fraud detection and credit risk applications are particularly well suited to such kind of analysis. This attitude commonly employs decision tree or neural network-based classification algorithms. Mainly data classification process involves learning and classification. In learning the training data are analysed by classification algorithm. During classification test data are used to estimate the accuracy of the classification rules. In the case of accuracy is acceptable the rules can be applied to the new data tuples. For a fraud detection type of application, this would comprise complete records of both fraudulent and valid activities determined on a record-by-record basis. To determine the set of parameters required for appropriate discrimination the classifier-training algorithm uses these pre-classified examples. Then the algorithm encodes these parameters into a model called a classifier.

Types of classification models:
1. Classification by decision tree induction
2. Bayesian Classification

3. Neural Networks
4. Classification Based on Associations

4.2 Artificial Neural Networks

An ANN also called as neural network is a mathematical model based on biological neural networks. Artificial neural network is based on observation of a human brain [3].Human brain is very complicated web of neurons. Analogically artificial work is an interconnected set of three simple units namely input, hidden and output unit. The attributes that are passed as input to the next form a first layer. In medical diagnosis patients risk factors are treated as input to the artificial neural network.
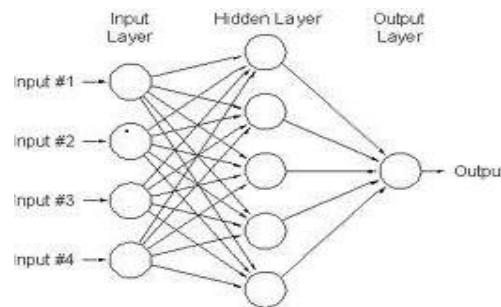

Fig 4.1: Artificial Neural Network (ANN)

The main function of artificial neural network is prediction. The effectiveness of artificial neural network was proven in medicine [2]. The noteworthy achievement of neural network was in the application of coronary heart disease.

There is numerous advantages of ANN some of these include
1) High Accuracy.
2) Independent from earlier assumptions about the distribution of the data.
3) Noise tolerance.
4) Ease of Maintenance.
5) ANN can be implemented in parallel hardware.

4.3 Data Mining Through Genetic Algorithms

We start out with a randomly selected first generation. Each string in this generation is assessed according to a fitness value is assigned and its quality. Then next, a new generation is created by applying the reproduction operator. After that pairs of strings of the new generation are selected and crossover is accomplished. With a certain probability, genes are mutated before all solutions are evaluated again. Then same procedure is followed till a maximum number of generations are reached. At the time of doing this, the all-time best solution is stored and returned at the end of the algorithm.

Genetic algorithm have been used in [4], to reduce the actual data size to get the optimal subset of attributed sufficient for heart disease prediction. Classification method is a supervised learning method to predict future trends or to extract models describing important data classes. Three classifiers e.g. Classification, Decision Tree, Naïve Bayes have been used to diagnose the presence of heart disease in patients. By considering with a certain possibility, genes are transformed before all solutions are evaluated again. Same procedure is repeated till a maximum number of generations are reached.

V. CONCLUSION

The design of the system requires the complete understanding of the problem domain. As per the above discussion there is need to efficiently diagnosis the presence of heart disease in an individual. The main objective of this review is to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease i.e Heart Attack using historical heart database. Originally 13 attributes were used for Heart Disease Prediction but the same work can also be achieved with the help of less number of attributes. So that the required efforts will get reduced. By considering patients basic information and other attributes value, it will be beneficial to improve the accuracy of existing algorithm approaches. It is also possible to classify the diagnosed results into four categories such as Healthy, Mild Attack, Moderate Attack And Severe Attack. More ever the results may also be stored in the database connected with the system for calculating the efficiency and for the record maintenance of the patients. A unique Patient ID can be generated by the system for each patient which may play important role in overall processing of the system. We are having several options for data

mining classification techniques namely Neural Networks, Decision Trees, and Naive Bayes can be used as Classifiers.

References

[1]   Sellappan P., Rafiah A., Intelligent Heart Disease Prediction System Using Data Mining   Techniques, IJCSNS, VOL.8, Aug 2008.
[2]   M. A. Jabbar , B.L Deekshatulu, Priti Chandra, Classification of Heart Disease Using Artificial Neural Network and Feature Subset Selection, GJCST, Volume13, Issue 3
[3]   Nang Y. Hand book of Data Mining, Lawrence Erlbaum associates (2003).
[4]   M. A. Jabbar et.al., Predictions Of Risk Score For Heart Disease Using Associative Classification & Hybrid Feature Subset Selection. In Proceedings of 12th International Conference on Intelligent Systems Design and Applications (ISDA), pages 628-634.
[5]   M. Ambarasi et.al, Enhanced Prediction of Heart Disease With Feature Subset Selection Using Genetic Algorithm, JEST Vol2, pp 5370-5376.
[6]   M. A. Jabbar et.al, Knowledge Discovery Using Associative Classification For Heart Disease Prediction, AISC 182 page no.29-39, Springer.
[7]   M. A. Jabbar et.al, Knowledge Discovery From Mining Association Rules For Heart Disease Prediction, JAJIT, Vol 41, pages 45-51.
[8]   M. A. Jabbar, B. Deekshatulu, Priti C., An Evolutionary Algorithm For Heart Disease Prediction, CCIS pages, 378-389springer Verlag
[9]   Sarojini B. et al, Feature Subset Selection Using FCBF In Type II Data Bases, ICIT Thailand March (2009).
[10]   Aqueel Ahmed, Shaikh Abdul Hannan, Data Mining Techniques To Find Out Heart Diseases: An Overview, IJITEE ISSN: 2278-3075, Volume-1, Issue-4, 2012.
[11]   Frawley and Piatetsky,1996. Knowledge Discovery in Databases An Overview. The AAAI/MIT Press, Menlo Park, C.A.
[12]   Miller, A., B. Blott and T. Hames, 1992. Review Of Neural Network Applications In Medical Imaging And Signal Processing. Med. Biol. Engg. Comp., 449-464.
[13]   Chen, J., Greiner, R.: Comparing Bayesian Network Classifiers, Proc. of UAI-99, Pages.101–108, 1999.
[14]   Srinivas, K., Analysis Of Coronary Heart Disease And Prediction Of Heart Attack In Coal Mining Regions Using Data Mining Techniques, IEEE Transaction on Computer Science and Education (ICCSE), pages(1344 - 1349), 2010.
[15]   http://heart-disease.emedtv.com/ coronary-artery-disease/coronary-artery-disease.html