A Study on Prediction of Cardiovascular Victimisation Data Processing Techniques

Dr.G.M.Nasira[#1], Mrs.S.Radhimeenakshi[*2],
[#1]Assistant Professor, Department of Computer Applications, Chikkanna Govt.college, Tirupur, Tamilnadu
[#2]Assistant Professor, Department of Computer Science, Tiruppur Kumaran College for Women, Tirupur, Tamilnadu

***Abstract*—** Heart disease term is related to several medical conditions of heart. Heart disease is one of the major health problems in India. The medical conditions refer to the abnormal health conditions that affect the heart. This paper presents a literature review of various data mining techniques implemented in prediction of heart disease. The observations reveal that neural networks and decision tree has more performance than all other data mining techniques.

***Keywords*—** Heart disease, Artificial Neural Network, Decision Tree, Naïve Bayes, Neuro-fuzzy, Genetic Algorithm, Association Rules.

Introduction

Data mining is used to discover the unknown knowledge from the known information and build predictive models. It is a step to discover knowledge from the data bases. This discovered knowledge can be utilized by the medical practitioners to reduce the time in diagnosis.

Nowadays medical business has complicated databases with the patient's records, disease, e-patient records, medical instruments, smarter treatment technique, and reduces insurance fraud, acknowledges high risk patients and hospital resources etc. These complex databases has to be analyzed and processed for getting knowledge to support cost-saving and decision-making.

Data mining comes with a set of tools and techniques which can be applied to the processed data to discover the unknown or hidden knowledge and support the medical practitioners with information for decision-making. The data mining may accomplish the functionalities like description, prediction, association, classification, clustering and time series analysis. The data mining has six phases which is given in figure 1.

Cardio Vascular Disease (CVD), the other name for heart disease has number of medical conditions that result in heart attacks. The problems such as heart valve abnormalities, decreased blood circulation and oxygen and also irregular heartbeats may also result in heart attack.

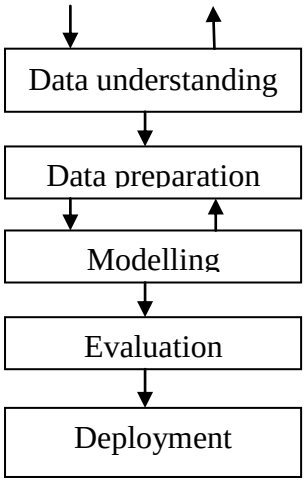

Figure 1.1: phases of Data Mining

Efficient and effective automated systems will be much more useful in medical industry. This work is attempted to review and analysis various data mining techniques that can be deployed in the automated systems. The automation will reduce the number of tests taken. So this saves time and cost of patient as well as medical practitioner.

2. METHODOLOGY

This paper exhibits the study of different data mining techniques which are helpful for medical practitioners for accurate prediction of heart disease. The methodology used for this work is by examining the journals, publications and reviews in the field of data mining bio-medicine and bio informatics, and heart disease.

3. DATA MINING AND PREDICTION OF HEART DISEASE

One of the significant challenges of data mining in health care is to obtain quality and relevant data. There are several data mining techniques that can be applied based on domains. Statistics provide a strong background for giving the result. The statistical algorithms must be modified and scaled before applying the data mining. Now we analyse some of the papers that have used various techniques of data mining. From the analysis it has been found out that most of them have taken the data from the Cleveland

database- UCI repository and have used the following attributes.

1. Age
2. Sex
3. Chest pain type (4 values)
4. Resting blood pressure
5. Serum cholesterol in mg/dl
6. Fasting blood sugar > 120 mg/dl
7. Resting electrocardiographic results (Values 0, 1, 2)
8. Maximum heart rate achieved
9. Exercise induced angina
10. Old peak = ST depression induced by Exercise Relative to rest
11. The slope of the peak exercise ST Segment
12. Number of major vessels (0-3)
13. Thal 3 = normal; 6 = fixed defect; 7 = Reverse defect

A. Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network (IEHAP) [1]

This paper is provided with an efficient approach to predict heart disease. The preprocessed data is then clustered with the use of k-means clustering algorithm and MAFIA algorithm for extracting the frequent items. These frequent items with high weightage are selected for prediction. The MLPNN is trained with the significant patterns using Back-propagation algorithm. This automated system is implemented in Java. It has used data from uci database and 13 attributes to predict the heart disease.

B. Intelligent Heart Disease Prediction Using Data Mining Techniques (IHDPS) [2]

This paper presents a prototype heart disease prediction system which is developed using three data mining classification modeling techniques: Decision trees, Naïve Bayes and Neural Network and implemented in .NET platform. DMX query language and functions are used to develop and access the models which are trained and validated against a test dataset. Lift chart and classification methods evaluate the models effectiveness. The highest effective model is Naïve Bayes, and next comes the neural network followed by decision tree.

C. Constraining and Summarizing Association Rules in Medical Data (CSARM) [3]

The aim of this paper is to discover the significant association rules and predict the search process. Constraints are proposed for filtering purposes to get significant rules. Greedy algorithm is used to summarize rules having same consequent. The significant rules are evaluated by three metrics: support, confidence and lift. Experiments have used real data set to predict the heart disease effectively. This proposed work can also be implemented in other domains.

D. *Intelligent Heart Disease Prediction System Using CANFIS and Genetic Algorithm (IHDPS-CG) [4]*

This paper presents a new approach for prediction of heart disease namely coactive neuro-fuzzy inference system (CANFIS). The proposed CANFIS model combines the adaptive capabilities of Neural Network and qualitative measures of fuzzy logic which then integrated with genetic algorithm to predict the heart disease. The author depicts that genetic algorithm is useful for auto-tuning of the CANFIS parameters and select feature set. This is proposed as an effective method to find non-linear relationship and mapping between the different data attributes.

E. Applications of Data Mining Techniques in HealthCare and Prediction of Heart Attacks (ADMHPH) [5]

This paper studied and focused on using different algorithms to predict combinations of several attributes. The authors in this work analyze the use of data mining techniques like Rule based, Decision Tree, Naïve Bayes and Neural Network. For preprocessing and decision making they used one dependency Augmented Naïve Bayes classifier (ODANB) and naïve creedal classifier 2. The attributes such as age, sex, blood pressure and blood sugar are used to predict the heart disease.

F. Mining Constrained Association rules to predict Heart Disease (MCAR) [6]

This work focuses on two aspects: mapping data to an association rule and identifying the constraints. Mapping algorithm transforms the medical records into transaction form. It treats the attributes as numerical or categorical. After doing these aspects the authors use several constraints to reduce the rules and speed the mining process.

G. Decision Support in Heart Disease Prediction system using Naïve Bayes (DSHDPS) [7]

This DSHDPS is developed using Naïve Bayes, a data mining technique. It is implemented as web based application in the form of questionnaire. Based on answers, it can identify and extract hidden knowledge related to heart disease. The report of the same is given to the patient in the form of pdf and chart. This model can answer complex queries, access information and prediction accuracy and incorporates categorical data.

H. Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers (IEHDPS) [8]

The main aim of this research work is to build an intelligent heart disease prediction system using weighted Associative classifiers and this is implemented in Java. This can be used in making

expert decision with high accuracy value. The advantage of this system is that it can be updated easily when the new training data set is available and moreover the prediction rules are already generated and stored in rule base. This again uses the UCI machine learning dataset which is used for research purpose.

4. THE TABLE SHOWING THE PROS AND CONS OF VARIOUS DATA MINING TECHNIQUES.

S.no	Paper	Advantages	Disadvantages
1	IEHAP [1]	Effective heart attack prediction by extracting patterns.	Large number of patterns so the running speed is increased
2	IHDPS [2]	Can answer complex queries, web-based, user-friendly, scalable, reliable and expandable.	Involves large number of rules, uses only categorical data, applied only for small data set
3	CSARM [3]	Reduce the number of discovered rule by constraints, improves running time by mapping information into binary format.	Constraints are specified on attributes but they are actually used on items. It is not possible to apply the standard association rule directly on a dataset with alphanumeric value.
4	IHDPS-CG [4]	New approach-CANFIS: combines neural network adaptive capabilities & fuzzy logic qualitative approach and integrated with genetic algorithm. Great potential in HDP by identifying non-linear relationship	Multiple generations (iterations) in the exploring the best network parameters
5	ADMHPH [5]	Delivers robust classifications when dealing with small or incomplete data sets	Specifies only five goals
6	MCAR [6]	Maps medical data to a transaction format and identifies useful constraints.	Does not deal with rule semantics, mapping rule size or noisy data. Uses only two transactions to mine medical data
7	DSHDPS [7]	Web based questionnaire application, reduces medical error, and integrates clinical decision support with computer based patient records.	Suitable particularly when the dimensionality of the inputs is high
8	IEHDPS [8]	GUI based interface and improved accuracy	Used with small data size

4. RESULT AND SUGGESTION

Different techniques of Data Mining are incorporated and researched by different authors. By analyzing these research papers it is clear that the medical data mining is rich in information but have to gain more knowledge.

- The papers have utilized statistical attributes but other methods like feature selection subset can be used to select attributes which in turn can enhance the performance and accuracy of the classification model.
- There is no single classifier that gives best result on the given data set. All the methods have used two sets- training and test data set. But there will be some problem in test data set. Sometimes it will be complex and sometimes easy depending upon the size of the test data set. To avoid this we can use cross validation method in which dataset is used for both training and testing.
- To avoid the problem of sampling biases there is no need to repeat the sampling process. Partitioned algorithm can be used to determine the number of clusters.
- While Classification rules discover the class of attributes but do not focus on the relationship among the attributes, Association focus upon the relationships.

5. CONCLUSION

This work concludes that no single data mining technique give consistent result. It depends on the data type we have taken for our work. So we can use hybrid or integrated data mining technique such as fusion of classification, fusion of clustering, association with classification or clustering etc. This work gives a study of different data mining techniques that are used in the prediction of heart disease. For an effective use of data mining techniques there should be a data sharing between the researcher and the hospitals to overcome the security issues. Each technique comes with its own consequences. Moreover the climatic conditions may have influence over the attribute values. So we have to strictly adhere to the local datasets.

REFERENCES

[1] Shantakumar B.Patil, Y.S.Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Neural Network", European Journal of Scientific Research (June 2009), 642-656.

[2] Sellappan Palaniappan, Rafiah Awang "Intelligent Heart Disease prediction System using Data Mining Techniques" IJCSNS Aug-2008.

[3] Carlos Ordonez, Norberto Ezquerra, Cesar A. Santana "Constraining and Summarizing Association Rules in Medical Data"Springer-verlag, 2006

[4] Latha Parthiban and R.Subramanian "Intelligent Heart Disease Prediction System using CANFIS and Genetic

Algorithm "International Journal of Biological and Life Sciences-2007.

[5] K.Srinivas, B.Kavihta Rani, Dr.A.Govrdhan "Applications of DataMining Techniques in Health care and Prediction of Heart Attacks" – IJCSE-VOL: 02, No.02, 2010

[6] Carlos Ordonez, Norberto Ezquerra, Cesar A. Santana "Mining Constrained Association Rules to Predict Heart Disease" IEEE 2001-International conference on data mining, 433-440

[7] G.Subbalakshmi, K.Ramesh, M.Chinna Rao "Decision Support in Heart Disease Prediction System Using Naive Bayes" IJCSE –Apr-May 2011.

[8] Jyothi Soni, Uzma Ansari, Dipesh Sharma, Sunitha Soni "Intelligent and Effective Heart Disease Prediction System Using Weighted Associative Classifiers" IJCSE- June 2011