IMPORTANCE OF CARDIO VASCULAR DATABASE IN MEDICAL DATAMINING

V. Geetha,
Assistant Professor,
Department Of Computer Science and Engineering,
SCSVMV University,
Kanchipuram,Tamilnadu,India.

T. Jayanthi,
Assistant Professor,
Department Of Computer Science and Engineering,
SCSVMV University,
Kanchipuram,Tamilnadu,India.

Abstract: Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. Some experts believe the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending. This could be a win/win overall. But due to the complexity of healthcare and a slower rate of technology adoption, our industry lags behind these others in implementing effective data mining and analytic strategies. Most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is a complex activity. This research intends to provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction.

Key words: Data mining, Decision Tree, Neural Network, Naive Bayes, cardiovascular disease.

I.INTRODUCTION

Data mining [1] can mean different things to different people. The most basic definition of data mining is the analysis of large data sets to discover patterns and use those patterns to forecast or predict the likelihood of future events. That said, not all analyses of large quantities of data constitute data mining. Generally categorize analytics as follows:

- Descriptive analytics—Describing what has happened
- Predictive analytics—Predicting what will happen
- Prescriptive analytics—Determining what to do about it

Data mining has been used to uncover patterns from the large amount of stored information and then used to build predictive models. While other solutions might favor healthcare providers or insurance companies, data mining benefits everyone concerned, from healthcare organizations to insurers to patients. Data mining has become a fundamental methodology for computing applications in medical informatics. Progress in data mining applications [2] and its implications are manifested in the areas of information management in healthcare organizations, health informatics, epidemiology, patient care and monitoring systems, assistive technology, large-scale image analysis to information extraction and automatic identification of unknown classes. Various algorithms associated with data mining have significantly helped to understand medical data more clearly, by distinguishing pathological data from normal data, for supporting decision-making as well as visualization and identification of hidden complex relationships between diagnostic features of different patient groups.

II. DATA MINING ALGORITHMS USED FOR PREDICTIONS

The datamining algorithms [3]are Decision tree, Neural Networks, Association algorithms, classification algorithms and Naïve bayes.

2.1. Neural Networks

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural network. In other words, it is an emulation of biological neural system [3]. A Multi-layer Perceptron Neural Networks (MLPNN) is used.

It maps a set of input data onto a set of appropriate output data. It consists of 3 layers input layer, hidden layer & output layer. There is connection between each layer & weights are assigned to each connection. The primary function of neurons of input layer is to divide input xi into neurons in hidden layer. Neuron of hidden layer adds input signal xi with weights wji of respective connections from input layer. The output Yj is function of $Yj = f(\Sigma\ wji\ xi)$ Where f is a simple threshold function such as sigmoid or hyperbolic tangent function.

2.2. Decision Trees

The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree & applying the tree to the dataset. There are many popular decision tree algorithms CART, ID3, C4.5, CHAID, and J48. From these J48 algorithm is used for this system. J48 algorithm uses pruning method to build a tree. Pruning is a technique that reduces size of tree by removing over fitting data, which leads to poor accuracy in predications. The J48 algorithm recursively classifies data

until it has been categorized as perfectly as possible. This technique gives maximum accuracy on training data. The overall concept is to build a tree that provides balance of flexibility & accuracy.

2.3. Naive Bayes

Naive Bayes classifier is based on Bayes theorem. This classifier algorithm uses conditional independence, means it assumes that an attribute value on a given class is independent of the values of other attributes.

The Bayes theorem is as follows:

Let X={x1, x2, ...,xn} be a set of n attributes.

In Bayesian, X is considered as evidence and H be some hypothesis means, the data of X belongs to specific

We have to determine P (H|X), the probability that the hypothesis H holds given evidence i.e. data sample X. According to Bayes theorem the P (H|X) is expressed as

$$P(H|X) = P(X|H)\,P(H) / P(X)$$

2.4. Associative Classification

Associative classification is a recent and rewarding technique which integrates association rule mining and classification to a model for prediction and achieves maximum accuracy. Associative classifiers are especially fit to applications where maximum accuracy is desired to a model for prediction.

Association rule mining and classification are two main functionalities of data mining. Association rule mining is used to find associations or correlations among the item sets. It is a unsupervised learning where no class attribute is involved in finding the association rule. On the other hand, classification is a supervised learning where class attribute is involved in the construction of the classifier and is used to classify or predict the data unknown sample.

III. IMPORTANCE OF COMPUTATIONAL ALGORITHMS IN CARDIOVASCULAR DISEASE DATABASE

The term Heart disease [4] encompasses the diverse diseases that affect the heart. Heart disease kills one person every 34 seconds around the world. Coronary heart disease, Cardiomyopathy and Cardiovascular disease are some categories of heart diseases. The term "cardiovascular disease" [5] includes a wide range of conditions that affect the heart and the blood vessels and the manner in which blood is pumped and circulated through the body. Cardiovascular disease (CVD) [6,7] results in severe illness, disability, and death. Narrowing of the coronary arteries results in the reduction of blood and oxygen supply to the heart and leads to the Coronary heart disease (CHD). A sudden blockage of a coronary artery, generally due to a blood clot results in a heart attack. Chest pains arise when the blood received by the heart muscles is inadequate and inductive logic programming. Many healthcare organizations struggle with the utilization of data collected through an organization online transaction processing (OLTP) system that is not integrated for decision making and pattern analysis. For successful healthcare organization it is important to empower the management and staff with data warehousing based on critical thinking and knowledge management tools for strategic decision making. Data warehousing can be supported by decision support tools such as data mart, OLAP and data mining tools. A data mart is a subset of data warehouse. It focuses on selected subjects. Online analytical processing (OLAP) solution provides a multi-dimensional view of the data found in relational databases. With stored data in two-dimensional format OLAP makes it possible to analyze potentially large amount of data with very fast response times and provides the ability for users to go through the data and drill down or roll up through various dimensions as defined by the data structure. So there is a need of cardiovascular database to predict this disease as earlier as possible to save the life of patients.

IV. DATA MINING APPLICATIONS IN HEALTHCARE SECTOR

Healthcare industry today generates large amounts of complex data about patients, hospital resources, disease diagnosis, electronic patient records, medical devices etc. Larger amounts of data are a key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Data mining applications in healthcare can be grouped as the evaluation into broad categories

Treatment effectiveness : Data mining applications can develop to evaluate the effectiveness of medical treatments. Data mining can deliver an analysis of which course of action proves effective by comparing and contrasting causes, symptoms, and courses of treatments.

Healthcare management : Data mining applications can be developed to better identify and track chronic disease states and high-risk patients, design appropriate interventions, and reduce the number of hospital admissions and claims to aid healthcare management. Data mining used to analyze massive volumes of data and statistics to search for patterns that might indicate an attack by bio-terrorists.

Customer relationship management : Customer relationship management is a core approach to managing interactions between commercial organizations-typically banks and retailers-and their customers, it is no less important in a healthcare context. Customer interactions may occur through call centers, physicians' offices, billing departments, inpatient settings, and ambulatory care settings.

Fraud and abuse : Detect fraud and abuses establish norms and then identify unusual or abnormal patterns of claims by physicians, clinics, or others attempt in data mining applications. Data mining applications fraud and abuse applications can highlight inappropriate prescriptions or referrals and fraudulent insurance and medical claims.

Medical Device Industry : Healthcare system's one important point is medical device. For best communication work this one is mostly used. Mobile communications and low-cost of wireless bio-sensors have paved the way for development of mobile healthcare applications that supply a convenient, safe and constant way of monitoring of vital signs of patients[3]. Ubiquitous Data Stream Mining

(UDM) techniques such as light weight, one-pass data stream mining algorithms can perform real-time analysis on-board small/mobile devices while considering available resources such as battery charge and available memory.

V. CONCLUSION

This paper aimed to compare the different data mining application in the healthcare sector for extracting useful information. The prediction of diseases using Data Mining applications is a challenging task but it drastically reduces the human effort and increases the diagnostic accuracy. Developing efficient data mining tools for an application could reduce the cost and time constraint in terms of human resources and expertise. Exploring knowledge from the medical data is such a risk task as the data found are noisy, irrelevant and massive too. In this scenario, data mining tools come in handy in exploring of knowledge of the medical data and it is quite interesting. It is observed from this study that a combination of more than one data mining techniques than a single technique for diagnosing or predicting diseases in healthcare sector could yield more promising results. The comparison study shows the interesting results that data mining techniques in all the health care applications give a more encouraging level of accuracy.. The overall objective is to study the various data mining techniques available to predict the heart disease and to compare them to find the best method of prediction.

VI. REFERENCES

[1] HianChyeKoh and Gerald Tan, "Data Mining Applications in Healthcare", journal of Healthcare Information Management – Vol 19, No 2.

[2] HeonGyu Lee, Ki Yong Noh, KeunHoRyu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.

[3] ShantakumarB.Patil, Y.S.Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network". ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656.

[4] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.

[5] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and KeunHoRyu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006.

[6] JayanthiRanjan, "Applications of data mining Techniques in pharmaceutical industry", Journal of Theoretical and Applied Technology, (2007).

[7] Niti Guru, Anil Dahiya, NavinRajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business R eview, Vol. 8, No. 1 (January - June 2007).