

A Detailed Review of Data Mining Techniques in Healthcare Management Systems

S. Nagendra¹ S. Punithamary² R. Somasundara Manikandan³

¹Assistant Professor & Head ^{2,3}Assistant Professor

^{1,2,3}Department of Computer Engineering

^{1,2}Idhaya College for Women, Sarugani, India ³Raja Doraisingam Govt. Arts College, Sivaganga, India

Abstract— In the recent trends of the healthcare industry, data mining techniques play a significant role in prediction and diagnosis of the diseases. Generally, healthcare organizations generate and collect a large volume of data in patient details. So the increment of data volume automatically requires the data to be retrieved when the person needed. Usually, there are large numbers of data mining applications are found in the healthcare industry, such as medical device management systems, the pharmaceutical industry, and Hospital information management systems. The main purpose of the data mining applications in the healthcare industry is to find the useful and hidden knowledge from the database. This paper provides a detailed review of different data mining applications in the healthcare industry and to reduce the complexity of the study of the healthcare data transactions. This paper also gives a detailed comparative analysis of various data mining techniques which are concerned for extracting knowledge from the databases generate in the healthcare industry.

Key words: Data Mining, KDD Process, Big Data, Healthcare Management Systems

I. INTRODUCTION

The main purpose of the data mining is to extract useful information from large databases or data warehouses. Data mining applications are used for both commercial and scientific sides. This research work fully focused the Data Mining applications on the scientific side. Scientific data mining distinguishes itself in the sense that the nature of the datasets is often very different from traditional market-driven data mining applications [3]. In this research work, a detailed survey is carried out on data mining applications in the healthcare sector, types of data used and details of the information extracted. Data mining algorithms applied in healthcare industry play a significant role in prediction and diagnosis of the diseases. To find the useful and hidden knowledge from the database is the purpose behind the application of data mining. Popularly data mining called knowledge discovery from the data. The knowledge discovery is an interactive process, consisting of developing an understanding of the application domain, selecting and creating a data set, preprocessing, data transformation. Data Mining has been used in a variety of applications such as marketing, customer relationship management, engineering, and medicine analysis, expert prediction, web mining and mobile and mobile computing.

In healthcare institutions leak the appropriate information systems to produce reliable reports with respect to other information in purely financial and volume related statements [1]. Data mining tools to answer the question that traditionally was a time to consuming and too complex to resolve. They prepare databases for finding predictive

information. Data mining tasks are Association Rule, Patterns, Classification and Prediction, Clustering. Most common modeling objectives are classification and prediction [2]. The reason that attracted a great deal of attention in information technology for the discovery of useful information from large collections is due to the perception that we are data rich but information poor. Some of the sample data mining applications are:

- Developing models to detect fraudulent phone or credit-card activity
- Predicting good and poor sales prospectus.
- Predicting whether a heart attack is likely to recur among those with cardiac disease.
- Identifying factors that lead to defects in a manufacturing process.

Expanding the health coverage to as many people as possible, and providing financial assistance to help those with lower incomes purchase coverage. Eliminating current health disparities would decrease the costs associated with the increased disease burden borne by certain population groups. Health administration or healthcare administration is the field relating to leadership, management, and administration of hospitals, hospital networks, and health care systems [4].

The structure of the paper is formatted as: Section 1 gives a detailed introduction related to data mining applications in the healthcare sector. Section 2 deals the related works to data mining applications. Section 3 discusses the design methodology of the proposed system. Section 4 deals the results & discussions with the comparative analysis of the Data Mining Applications in Health Care Sector. The conclusion is defined in the next section 5.

II. RELATED WORK

There are many works have been done so far in the field of the data mining techniques in healthcare management systems. The following papers are the recent works related to the healthcare management systems in data mining techniques.

In 2011, HianChyeKoh and Gerald Tan [5] discusses data mining and its applications with major areas like Treatment effectiveness, Management of healthcare, Detection of fraud and abuse, Customer relationship management.

In 2014, Jayanthi Ranjan [7] presents how data mining discovers and extracts useful patterns of this large data to find observable patterns. This paper demonstrates the ability of Data mining in improving the quality of the decision-making process in the pharmacy industry. Issues in the pharma industry are adverse reactions to the drugs.

In 2013, M. Durairaj, K. Meena [4] illustrates a hybrid prediction system consists of Rough Set Theory (RST) and Artificial Neural Network (ANN) for dispensation

medical data. The process of developing a new data mining technique and software to assist competent solutions for medical data analysis has been explained. Propose a hybrid tool that incorporates RST and ANN to make proficient data analysis and indicative predictions. The experiments on spermatological data set for predicting excellence of animal semen is carried out. The projected hybrid prediction system is applied for pre-processing of the medical database and to train the ANN for production prediction. The prediction accuracy is observed by comparing observed and predicted cleavage rate.

In 2011, K. Srinivas, B. Kavitha Rani and Dr. A. Goverdhan [6] discusses mainly to examine the potential use of classification based data mining techniques such as Rule-Based, Decision tree, Naïve Bayes and Artificial Neural Network to the massive volume of healthcare data. Using an age, sex, blood pressure and blood sugar medical profiles it can predict the likelihood of patients getting a heart disease. In 2013, Shweta Kharya [5] discussed various data mining approaches that have been utilized for breast cancer diagnosis and prognosis Decision tree is found to be the best predictor with 93.62% Accuracy on benchmark dataset and also on SEER data set.

III. DATA MINING APPLICATIONS IN HEALTH CARE 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 [5],

A. 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.

B. 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 bioterrorists.

C. 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.

D. 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.

E. Medical Device Industry

Healthcare system's one important point is the 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 the development of mobile healthcare applications that supply a convenient, safe and constant way of monitoring of vital signs of patients [7]. Ubiquitous Data Stream Mining (UDM) techniques such as lightweight, one-pass data stream mining algorithms can perform real-time analysis onboard small/mobile devices while considering available resources such as battery charge and available memory.

F. Pharmaceutical Industry

The technology is being used to help the pharmaceutical firms manage their inventories and to develop new product and services. A deep understanding of the knowledge hidden in the Pharma data is vital to a firm's competitive position and organizational decision-making.

G. Hospital Management

Organizations including modern hospitals are capable of generating and collecting a huge amount of data. Application of data mining to data stored in a hospital information system in which temporal behavior of global hospital activities is visualized [1]. Three layers of hospital management:

- Services for hospital management
- Services for medical staff
- Services for patients

H. System Biology

Biological databases contain a wide variety of data types, often with rich relational structure. Consequently, multi-relational data mining techniques are frequently applied to biological data. Systems biology is at least as demanding as, and perhaps more demanding than, the genomic challenge that has fired international science and gained public attention.

IV. RESULTS ANALYSIS FOR DATA MINING APPLICATIONS IN HEALTH CARE SECTOR

In this chapter, a comparative study of data mining applications in the healthcare sector by different researchers gave in detail. Mainly data mining tools are used to predict the successful results from the data recorded on healthcare problems. Different data mining tools are used to predict the accuracy level in different healthcare problems. In this study, the following list of medical problems has been analyzed and evaluated.

- Heart Disease
- Cancer
- HIV/AIDS
- Blood
- Brain Cancer
- Tuberculosis

- Diabetes Mellitus
- Kidney dialysis
- Dengue
- Hepatitis C

In Table 1, the most important healthcare problems specifically in disease side and research results have been

illustrated. The diseases are the most critical problems inhuman. To analyze the effectiveness of the data mining applications for diagnosing the disease, the traditional methods of mathematical/statistical applications are also given and compared. Listed eleven problems are taken for comparison with this work.

S.No	Types of Disease	Data Mining Tool	Technique	Algorithm	Traditional Method	Accuracy Level (%) from DM Applications
1	Heart Disease	ODND	Classification	Naïve	Probability	60
2	Cancer	WEKA	Classification	Rules, Decision Table		97
3	HIV/AIDS	WEKA 3.6	Classification, Association Rule Mining	J48	Statistics	81
4	Blood Bank sector	WEKA	Classification	J48		89
5	Brain	K-Means Clustering	Clustering	MAFIA		85
6	Tuberculosis	WEKA	Naïve Bayes Classifier	KNN	Probability, Statistics	78
7	Diabetes Mellitus	ANN	Classification	C4.5 algorithm	Neural Network	82
8	Kidney Dialysis	RST	Classification	Decision Making	Statistics	75
9	Dengue	SPSS Modeler		C5.0	Statistics	80
10	Hepatitis	SNP	Information Gain	Decision Rule		73

Table 1: Analysis of Data Mining Applications in Health Care

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 risky 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 healthcare applications.

REFERENCES

[1] Arvind Sharma and P.C. Gupta, "Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool", International Journal of Communication and Computer Technologies, Volume 01 – No.6, Issue, September 2012.

[2] PrasannaDesikan, Kuo-Wei Hsu, JaideepSrivastava, "Data Mining for Healthcare Management", 2011 SIAM International Conference on Data Mining, April, 2011.

[3] B.Renuka Devi, Dr.K.NageswaraRao et al., "Disaster Prediction System Using IBM SPSS Data Mining Tool, International Journal of Engineering Trends and Technology (IJETT)", Volume4 Issue8, August 2013.

[4] Shouman, M., Turner, T., et al., "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients", International Conference on Knowledge Discovery (ICKD 2012).

[5] ShwetaKharya, "Using Data Mining Techniques For Diagnosis And Prognosis Of Cancer Disease", International Journal of Computer Science, Engineering and Information Technology (IJCSIEIT), Vol.2, No.2, April 2012.

[6] K. Srinivas, B. Kavitha Rani and Dr. A. Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering, 2010.

[7] Wen-Tsann Lin, Shen-Tsu Wang et al., "Abnormal diagnosis of Emergency Department triage explored with data mining technology: An Emergency", Expert Systems with Applications ISSN: 2733–2741, 2010.