A SURVEY ON DATA MINING TECHNIQUES USED IN HEALTH DIAGNOSIS

Shruthi P

Assistant Professor,
Department of Computer Science and Engineering,
ATME College of Engineering, Mysuru, India

Abstract : Nowadays health care centres is producing enormous quantity of data that includes complex information relating to patients and their medical conditions. Data mining is gaining popularity in different research areas due to its enormous applications and methodologies to mine the required information in technical and correct manner. Data mining techniques can be efficiently used to discover hidden patterns or relationships among the objects in the medical data. In last decade, the usage of data mining techniques on medical data for determining useful trends or patterns that are used in analysis and decision making has been increased drastically. Data mining has an infinite potential to utilize healthcare data more efficiently and effectually to predict different kind of disease. This paper features various applications of data mining techniques in medical healthcare and identifying appropriate data mining technique suitable in diagnosing different king of diseases.

Key words- Data Mining, Health Care, Data Mining Technique, Hidden Patterns.

I. INTRODUCTION

Data Mining is one of the most interesting and motivating area of research. The main objective of Data Mining is finding meaningful information from huge data sets. Data Mining has vital applications in healthcare field because there is a need of efficient analytical methodology for detecting unknown and valuable information in health data. In healthcare, Data Mining provides several advantages such as fraud detection in health insurance, availability of medical solution to the patients at lower cost, detection of causes of diseases and identification of medical treatment methods. It also benefits the healthcare researchers for making efficient healthcare policies, developing automated drug recommendation systems, creating health profiles of individuals etc.

Knowledge Discovery (KDD) and Data Mining are related terms and are used interchangeably but several researchers assume that both terms are dissimilar as Data Mining is one of the most vital stages of the KDD process. According to Fayyad et al., the Knowledge Discovery in database is implemented by using various stages whereas the first stage is selection of data in which data is collected from different sources, the second stage is pre-processing the selected data, the third stage is transforming the data into suitable format so that it can be processed further, the fourth stage consist of Data Mining where suitable Data Mining technique is applied on the transformed data for extracting valuable information and evaluation is the last stage.

Knowledge Discovery (KDD) process with the different stages involved is shown in Figure 1.
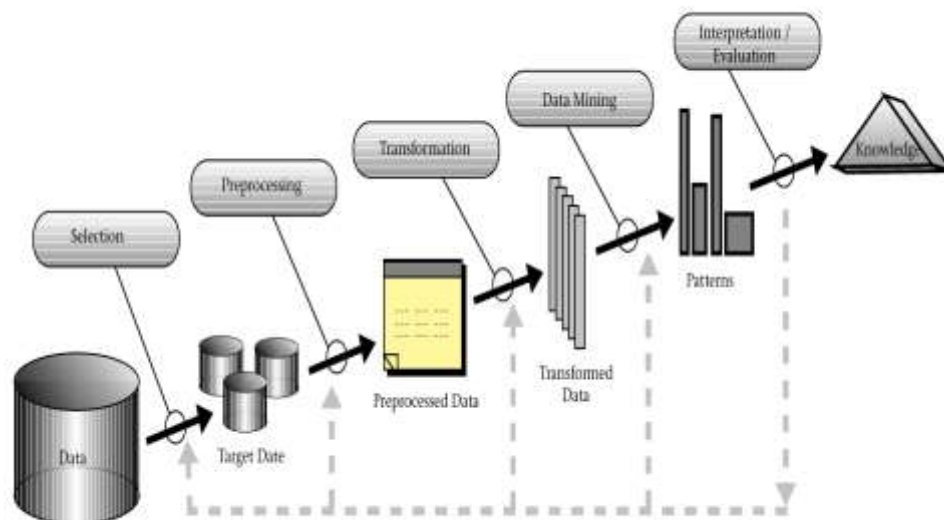

Figure 1: Knowledge Discovery in Databases

- **Selection step** In this step the heterogeneous data is collected from different sources for processing. Real time data collected from different healthcare may be incomplete, complex, noisy, inconsistent, and/or irrelevant. Hence the selection process selects the important data from which knowledge is to be extracted.
- **Pre-processing step** This step performs basic operations on the selected data for removing the noisy data, try to identify the missing data or to identify the methodology for handling missing data, identify or eliminate outliers and resolve inconsistencies among the data.

- **Transformation step** In this stage the pre-processed data is transformed into different forms which is suitable for mining the required information by performing task like aggregation, smoothing, normalization, generalization, and discretization. Data reduction task shrinks the huge amount of data and represents the same data in less volume, but while processing the data it will give the same analytical outcome as before.

II. RELATED WORK

1. The Survey of Data Mining Applications And Feature Scope

This paper have focused on variety of techniques, approaches and different areas of the research which are helpful and marked as the important field of data mining Technologies. As we are aware that many **MNC's** and large organizations are operated in different places of the different countries**.** Each place of operation may generate large volumes of data. Corporate decision makers require access from all such sources and take strategic decisions .The data warehouse is used in the significant business value by improving the effectiveness of managerial decision-making. In an uncertain and highly competitive business environment, the value of strategic information systems such as these are easily recognized however in today's business environment, efficiency or speed is not the only key for competitiveness. This type of huge amount of data's are available in the form of tera- to peta-bytes which has drastically changed in the areas of science and engineering. To analyze, manage and make a decision of such type of huge amount of data we need techniques called the data mining which will transforming in many fields. This paper imparts more number of applications of the data mining and also focuses scope of the data mining which will helpful in the further research.

2. Data Mining Techniques using in Medical Science

This article explores data mining techniques in health care. In particular, it discusses data mining and its application in areas where people are affected severely by using the under-ground drinking water which consist of high levels of fluoride. This paper identifies the risk factors associated with the high level of fluoride content in water, using clustering algorithms and finds meaningful hidden patterns which gives meaningful decision making to this socio-economic real world health hazard.

3. A survey on Data Mining approaches for Healthcare

Data Mining is one of the most motivating area of research that is become increasingly popular in health organization. Data Mining plays an important role for uncovering new trends in healthcare organization which in turn helpful for all the parties associated with this field. This survey explores the utility of various Data Mining techniques such as classification, clustering, association, regression in health domain. In this paper, we present a brief introduction of these techniques and their advantages and disadvantages. This survey also highlights applications, challenges and future issues of Data Mining in healthcare. Recommendation regarding the suitable choice of available Data Mining technique is also discussed in this paper.

4. Survey Of Data Mining Techniques Used In Healthcare Domain

patients and their medical conditions. Data mining is gaining popularity in different research arenas due to its infinite applications and methodologies to mine the information in correct manner. Data mining techniques have the capabilities to discover hidden patterns or relationships among the objects in the medical data. In last decade, there has been increase in usage of data mining techniques on medical data for determining useful trends or patterns that are used in analysis and decision making. Data mining has an infinite potential to utilize healthcare data more efficiently and effectually to predict different kind of disease. This paper features various Data Mining techniques such as classification, clustering, association and also highlights related work to analyse and predict human disease.

5. Critical Analysis of Data Mining Techniques on Medical Data

The use of Data mining techniques on medical data is dramatically soar for determining helpful things which are used in decision making and identification. The most extensive data mining techniques which are used in healthcare domain are, classification, clustering, regression, association rule mining, classification and regression tree (CART). The suitable use of data mining algorithm can enhance the quality of prediction, diagnosis and disease classification. Valuation of data mining techniques demand for medical data mining is the major goal here, particularly to examine the local frequent disease like heart ailments, breast cancer, lung cancer and so on. We examine for discovering the locally frequent patterns through data mining technique in terms of cost performance speed and accuracy.

III. APPLICATIONS OF DATA MINING TECHNIQUES IN HEALTH CARE

Nowadays the medical healthcare is generating large amounts of complex data about patients, their past medical history, hospital resources, disease diagnosis information, electronic patient records etc. These large amounts of data are the key resources which need to be processed and analysed for knowledge extraction that also helps in cost-savings and decision making. Data mining applications in healthcare can be categorized into different group of clusters.

1. Effective Treatment

Data mining applications can be used to identify the effectiveness level of medical treatments. Data mining can deliver an analysis of which course of action should be used to improve the treatment effectiveness by comparing and contrasting causes, symptoms, and courses of treatments.

2. Healthcare management

Data mining applications can be developed better identifying and tracking chronic disease stages and high-risk patients, design appropriate intrusion, and lessen the number of patients getting admitted to hospital and claims to aid healthcare management. Data mining used to analyze huge volumes of data and information to search for patterns that might indicate an attack by bio-terrorists.

3. Customer relationship management

Customer relationship management is a core approach for managing interactions between business organizations-typically banks and retailers-and their customers, it is also equally important in a healthcare centres. Customer interactions may occur through call centres, physicians' offices, billing departments, inpatient settings, and ambulatory care settings etc.

4. Fraud and abuse

Fraud and abuses detection establishes the norms and then categorize strange or anomalous patterns of claims by physicians, clinics, or others attempt in data mining applications. Data mining applications fraud and abuse detection can highlight wrong prescriptions or referrals and fake insurance and medical claims.

5. Medical Device engineering

Medical device is very important for healthcare centres. These devices are mainly used for best communication work. Mobile communications and low-cost of wireless bio-sensors have paved the way for development of mobile healthcare applications that provide a suitable, secure and stable way of monitoring the fundamental signs of patients. 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.

6. Pharmaceutical business

Data Mining techniques is being used to help the pharmaceutical firms in managing their inventories and to expand their business by developing new product and services. A deep thoughtful of the knowledge hidden in the Pharma data is very important to a firm's competitive position and organizational decision-making.

7. 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. Three layers of hospital management:

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

IV. SUMMARY OF TECHNIQUES FOR MEDICAL DATA MINING

The Data Mining techniques such as classification, prediction and decision making have substantial expansion in medical industry with respect to various diseases like diabetes, heart disease, liver diseases, cancer and others. Table 1 summarizes the medical data mining, its techniques, tools and appropriate algorithm usage. It also shows the comparison with traditional method and Data Mining method by measuring the accuracy level when Data Mining technique is used.

Table 1: Summary of Techniques for Medical Data Mining

Sl. No.	Types of Disease	Data Mining tool	Technique	Algorithm	Traditional Method	Accuracy level (%) from DM application
1	Heart Disease	ODND, NCC2	Classification	Naive	Probability	60
2	Cancer	WEKA	Classification	Rules, Decision Table		97.77
3	HIV/AIDS	WEKA 3.6	Classification, Association Rule Mining	J48	Statistics	81.8
4	Blood Bank Sector	WEKA	Classification	J48		89.9
5	Brain Cancer	K-means Clustering	Clustering	MAFIA		85
6	Tuberculosis	WEKA	Naive Bayes Classification	KNN	Probability, Statistics	78
7	Diabetes Mellitus	ANN	Classification	C4.5 algorithm	Neural Network	82.6
8	Kidney dialysis	RST	Classification	Decision Making	Statistics	75.97
9	Dengue	SPSS Modeller		C5.0	Statistics	80
10	IVF	ANN, RST	Classification			91
11	Hepatitis C	SNP	Information Gain	Decision rule		73.20

V. CONCLUSION AND FUTURE WORK

With the recent rapid rise in the quantity of biomedical data that is gathered by electronic means in critical care and the rampant availability of inexpensive and dependable computing equipment, many researchers has started, or are eager to start, exploring these data. In this paper we observe some data mining techniques that has been employed for medical data. As there is voluminous records in this industry and because of this, it has become requisite to use data mining techniques to help in decision support and prediction in the field of Healthcare to identify the kind of disease. The medical data mining produces business intelligence which is useful for diagnosing of the disease. This paper throws light into data mining techniques that is used for medical data for various diseases which are identified and diagnosed for human health.

REFERENCES

[1] **"The Survey of Data Mining Applications And Feature Scope"**, Neelamadhab Padhy, Dr. Pragnyaban Mishra, and Rasmita Panigrahi, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June 2012.

[2] **"Data Mining Techniques using in Medical Science"**, Potu Narayana, Sangam Venkata Ramana, A.Soujanya, International Journal of Electrical, Electronics and Computer Systems (IJEECS), ISSN (Online): 2347-2820, Volume -3, Issue-11 2015.

[3] **"A survey on Data Mining approaches for Healthcare"**, Divya Tomar and Sonali Agarwal, International Journal of Bio-Science and Bio-Technology Vol.5, No.5 (2013), pp. 241-266.

[4] **"Survey of Data Mining Techniques Used in Healthcare Domain"**, Sheenal Patel and Hardik Patel, International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016.

[5] **"Data Mining Applications In Healthcare Sector: A Study"**, M. Durairaj, V. Ranjani, International Journal of Scientific & Technology Research Volume 2, Issue 10, October 2013.

[6] **"Critical Analysis of Data Mining Techniques on Medical Data"**, Zahid Ullah, Muhammad Fayaz, Asif Iqbal, Published Online February 2016 in MECS (http://www.mecs-press.org/).

[7] **"A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases"**, Mohammed Abdul Khaleel, Sateesh Kumar Pradham, G.N. Dash, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013.

[8] **"Student Performance Prediction in Education Sector Using Data Mining"**, Shruthi P, Chaitra B P, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016.

BIOGRAPHY

Miss Shruthi P is working as an Assistant Professor in Department of Computer Science and Engineering, ATME College of Engineering, Mysuru, Karnataka, India. She received Master's in Computer Engineering (CE) in 2016 from PES College of Engineering, Mandya, Karnataka, India.