International Journal of Multidisciplinary
Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data Mining in Healthcare

Meghana H M[1], Sunith Kumar T[2]

Student, Department of MCA, St. Joseph Engineering College, Mangalore, Karnataka, India[1]

Assistant Professor, St. Joseph Engineering College, Mangalore, Karnataka, India[2]

ABSTRACT: Data mining has emerged as a powerful tool in the field of healthcare, revolutionizing how vast volumes of patient data can be analysed and transformed into valuable insights. These abstract highlights the significance of data mining in the healthcare sector, exploring its potential to improve patient outcomes, optimize healthcare processes, and advance medical research. The healthcare industry generates a large amount of data, encompassing patient records, medical imaging, electronic health records, clinical trials and more. Data mining techniques offer a solution by leveraging advanced algorithms to sift through complex datasets and discover valuable knowledge.

KEYWORDS: **Data transformation, Applications, Challenges**

I. INTRODUCTION

Data mining holds incredible potential for healthcare services due to the exponential growth in the number of electronic health records. Previously Doctors and physicians hold patient information in the paper where the data was quite difficult to hold. Digitalization and innovation of new techniques reduce human efforts and make data easily assessable. For example, the computer keeps a massive amount of patient data with accuracy, and it improves the quality of the whole data management system. Still, the major challenge is what should healthcare services providers do to filter all the data efficiently? This is the place where data mining has proven to be extremely useful [1,3].

II. DATA TRANSFORMATION

A. Health data sources: Health data can be collected from both hospitals and healthcare practitioners. Hospital data include patient's data, diagnosis, drug prescription and treatment data while healthcare practitioners' data are data collected from government healthcare agencies such was World Health Organization (WHO) and other healthcare.

B. organizations. The most reliable healthcare data comes from governmental sources or healthcare professional organizations. The data comes from several sources of varying quality, inconsistent and incorrect that are of immense volume. The volume of the data needs to be organized, structured and processed to get meaningful result. There are five most useful sources of data which are clinical data, claims data, patient-generated data and pharmaceutical data.

Healthcare data are stored in electronic format all over the world in health organization. The format of the data contains patient's details which are of vast data. Due to the increase in in data, there exist complexity and complications. It can be worrisome when using traditional methods in analysis this set of data to generate meaning knowledge from it. The field of mathematics, computer and statistics makes it easy to discover meaningful information from volumes of complex data which makes data mining to be of great benefits to the healthcare sector.

Data mining extracts meaningful information from complexity of data which were in a raw form. Numerous benefits are provided with the use of data mining in healthcare such as detection of fraud, detection of abuse of drugs, proper diagnosing of patients, treatments, early detection of diseases, survivability of patients etc.

Data mining techniques have been applied by various researchers. Such techniques are classification, association, clustering etc. the techniques play a vital role in the healthcare industry to support decision making, proper diagnosis, selection of treatments and prediction [3].

C. Health data transformation: data transformation is the process of changing data to information, usually from one format to another. The format from the source data is changed to the format of the desired information. Data needs to be transformed before it can make more sense. Data in a raw state is not meaningful and useful until it's transformed.

Different approach can be employed to transform data such as service oriented architectural and data warehousing. Data are made available in different form which are executed incrementally. Commonly used transformational languages are Perl, AWK, TEXT, XSLT and template languages and processors [1].

III. DATA MINING PROCESSES

The availability of the volume of data generated in healthcare industry need to be transform into a meaningful information for decision to occur. Data mining provides a great promise in analyzing complexity of data to generate information. The process of data mining helps to discover knowledge which are done in seven steps starting from selection stage to knowledge discovery. Selection Certain parameters are used to pick the data which is the first stage of the data mining. Preprocessing This stage nullify some of the parameter that are not needed. It helps to have a clean and correct data. Transformation It changes data that are necessary for specific solution. Data particular to that problem are transformed. Data mining this stage helps to discover knowledge from complexity of data. It is called a knowledge discovery stage. Interpretation and evaluation the information generated from data mining stage are evaluated. The evaluation of the data yield discovery knowledge from the complexity that will be useful for decision making.

IV. DATA MINING TOOLS USED IN HEALTHCARE

Data mining tools helps to analyzed volumes of complex data based on the dataset attributes that users specify in determining trends of occurrences. The software can be used for diagnoses, prediction, and management of diseases to extract knowledge and make decisions. Due to the availability of various software tools used, the choice of choosing appropriate software to solve a particular problem becomes difficult [24]. The most common data mining tools are explained below:

a.   WEKA (Waikato Environment for Knowledge Analysis): WEKA is a program tools used in data mining processes. It is a software that is develop using java programming language that runs on different operating system. Weka compliment several data mining processes. The software can connect with the data directly or from the java code. It make use of the Graphical user interface (GUI) for access to his performance and functionality.

b.   KEEL(Knowledge Extraction based on Evolutionary learning): KEEL uses clustering, regression, classification to extract pattern from datasets. It is an open source software but source program may be hidden. Complete analysis can be performed using the KEEL data mining tools.

c.   R: R is an open-source program for computation and statistical analysis. R software is of great benefits to the research and development world and health industry. The software for development of R data mining tools are FORTRAN, C and R.

d.   KNIME: Konstanz Information Miner, is an open-source software that are used for analyzing and modelling data. Machine leaning and data mining features are supported with KNIME software. KNIME has been applied in clinical research, detection of diseases and classification. KNIME can generate a work process that can documented in different format.

e.   RAPIDMINER: RAPIDMINER, data mining, machine learning, text mining and business analytics are development by an organization that provides the same software. It is used in business, finance, banking, insurance, medical, and education in analyzing data which support data mining processes. It is an open source software used in various filed of human endeavour.

f.   ORANGE: Orange is an open-source software used in machine learning. It is characterized by two features of front end and back end. The front end uses visual programming while back end uses python libraries. It was developed using C++ and Python programming. In science, new machine leaning algorithms and techniques in genetics, medical can be tested using ORANGE. It was used in education to teach student biology, science and other related courses to medical.

V. APPLICATIONS OF DATA MINING

Data mining has been used intensively and widely by numerous industries. In healthcare, data mining is becoming more popular nowadays. Data mining applications can incredibly benefit all parties who are involved in the healthcare industry. For example, data mining can help the healthcare industry in fraud detection and abuse, customer relationship management, effective patient care, and best practices, affordable healthcare services. The large amounts of data generated by healthcare transactions are too complex and huge to be processed and analysed by conventional methods [5].

Treatment effectiveness: Data Mining applications can be used to assess the effectiveness of medical treatments. Data mining can convey analysis of which course of action demonstrates effective by comparing and differentiating causes, symptoms, and courses of treatments [6].

Healthcare management: Data mining applications can be used to identify and track chronic illness states and incentive care unit patients, decrease the number of hospital admissions, and supports healthcare management. Data mining used to analyse massive data sets and statistics to search for patterns that may demonstrate an assault by bioterrorists.

Customer relationship management: Customer and management interactions are very crucial for any organization to achieve business goals. Customer relationship management is the primary approach to managing interactions between commercial organizations normally retail sectors and banks, with their customers. Similarly, it is important in the healthcare context. Customer interactions may happen through call centers, billing departments, and ambulatory care settings [6].

Fraud and abuse: Data mining fraud and abuse applications can focus on inappropriate or wrong prescriptions and fraud insurance and medical claims.

VI. CHALLENGES IN HEALTHCARE DATA MINING

One of the biggest issues in data mining in healthcare is that the raw medical data is huge and heterogeneous. These data can be accumulated from different sources [3]. For example, from conversations with patients, doctors review, and laboratory results. All these components can have a significant effect on the diagnosis, and treatment of a patient. Missing, incorrect, inconsistent data such as pieces of information saved in various formats from different data sources create a significant obstacle to successful data mining. Some of the key challenges in healthcare data mining are:

A. **Data Quality and Integration:** Healthcare data is often collected from various sources, including electronic health records, medical devices, insurance claims, and patient generated data. Ensuring data quality and integrating disparate datasets can be challenging, as data may be incomplete, inconsistent, or contain errors.
B. **Privacy and Security:** Healthcare data contains sensitive and personally identifiable information, making privacy and security paramount. Data mining techniques must adhere to strict regulations and guidelines to protect patient confidentiality and prevent unauthorized access or data breaches [3].
C. **Data Access and Sharing:** Data Sharing between healthcare institutions and researchers is essential for comprehensive analysis, but concerns about data ownership, legal constraints, patient, consent can impede collaboration and critical datasets.
D. **Lack of Standardization:** Healthcare data is often recorded in different formats and terminologies across institutions, hindering the seamless integration and analysis of data. Standardization efforts are necessary to enable meaningful comparisons and analysis.
E. **Imbalanced Data:** In healthcare, certain conditions or diseases may be relatively rare, leading to imbalanced datasets where some classes have significantly fewer instances. This can affect the performance of data mining algorithms, as they may be biased towards the majority class.

VII. CONCLUSION

The data mining has played in an important role in healthcare industry, especially in predicting various types of diseases. The diagnosis is widely being used in predicting diseases, they are extensively used in medical diagnosing and data mining has emerged as a transformative force in the healthcare industry,offering un precedented opportunities to revolutionize patient care, medical research, and operational efficiency. The vast amount of data generated in healthcare presents both challenges and opportunities, and data mining provides valuable tools to extract meaningful insights and Knowledge

REFERENCES

[1]   Canlas, R. D. "Data mining in healthcare: Current application and issues." School of Information Systems & Management, Carnegie Mellon University, Australia (2009).

[2]   Durairaj M, Ranjani V. Data mining applications in healthcare sector: a study. International journal of scientific & technology research. 2013 Oct 25;2(10):29-35.

[3]   Ahmad P, Qamar S, Rizvi SQ. Techniques of data mining in healthcare: a review. International Journal of Computer Applications. 2015 Jan 1;120(15).

[4]   Jothi, Neesha, and Wahidah Husain. "Data miningtin healthcare–a review." Procedia computer science 72 (2015): 306-313.

[5]   Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." International Journal of Bio-Science and Bio-Technology 5.5 (2013): 241-266.

[6]   Padhy, Neelamadhab, Dr Pragnyaban Mishra, and Rasmita Panigrahi. "The survey of data mining applications and feature scope." arXiv preprint arXiv:1211.5723 (2012).

[7]   Kalyani et al., International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 2, Issue 10, October 2012

[8]   Ogundele, I. O., O. L. Popoola, O. O. Oyesola, and K. T.   Oria. "A review on data mining in healthcare." International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) 7 (2018): 698-704.

[9]   Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." International Journal of Bio-Science and Bio-Technology 5.5 (2013): 241-266.

[10]   Padhy, Neelamadhab, Dr Pragnyaban Mishra, and Rasmita Panigrahi. "The survey of data mining applications and feature scope." arXiv preprint arXiv:1211.5723 (2012).

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY