

Mining Medical Data: A Comprehensive Study

Rama Krishna K^{1*}, K G Mohan²

^{1,2}Dept. of Computer Science, Presidency University, Bangalore, India

Corresponding Author: ramakrishna@presidencyuniversity.in

DOI: <https://doi.org/10.26438/ijcse/v7si16.98100> | Available online at: www.ijcseonline.org

Abstract- Health care industry generates huge volumes of data that comprises of complex information pertaining to patients’ data and their medical history. Data mining methodologies have the abilities to discover hidden patterns and associations among the attributes in the medical dataset. Due to Data Mining’s enormous applications, there has been an upsurge in usage of data mining techniques on medical data for discovering useful trends or knowledge patterns that are used in strategic decision-making process or disease diagnosis and treatment process. This paper focuses on need for mining medical data, issues and challenges surrounding Mining Medical Data.

Keywords- Data Mining, Health care, Knowledge Discovery (KDD), Decision-making

I. INTRODUCTION

In the yesteryears, it was highly expensive to store and maintain information. However, due to the progression in the area of information gathering and WWW in the last couple of decades, it is a fact that, vast volumes of information or data are obtainable in electronic format. To store such huge volumes of data or information the database size is enhanced briskly. Such databases comprises of highly valuable information. This information may be very beneficial for decision-making process irrespective of any field. Data mining is the procedure of digging out the beneficial information from a huge pool of data, which was formerly unknown [1]. A number of associations or relationships are concealed in such a huge set of data. For instance, an association/relationship involving patient’s data and their number of days of getting admitted [2].

In first phase, gathered raw data will be under mining, which will be converted into meaningful knowledge, or information shown after last phase in figure 1. In Data selection phase, data is selected rendering to certain norms. For instance, Data about the owners of cars. We can find subsets of data by making a selection of data. Data preprocessing phase eliminates the information, which is not required in final knowledge. For instance, data pertaining to the sex of the patient in not required when carrying out pregnancy test, so this field can be removed in patient’s dataset. Data Transformation phase, data is transformed or combined into forms suitable for mining by performing summary or aggregation tasks. In Data mining phase, intelligent techniques are applied in-order to extract useful and meaningful data patterns. Finally, the extracted knowledge will further be analyzed and fed into a visualization tool so that knowledge can be presented in the form of a graph or chart in-order to take effective decisions.

II. NEED FOR MINING MEDICAL DATA

In general, most of the healthcare firms across the globe stores patient’s data in electronic format. Healthcare information primarily consists of all the attributes of data regarding patients along with the parties associated in healthcare industries. Due to constant increase in the magnitude of electronic healthcare data, few complexities exist in it. In simple terms, data pertaining to healthcare has become increasingly complex. By adopting old-fashioned methods, it is very tough in to take out meaningful information from it. However, with the progression in fields like mathematics, statistics, and other disciplines, it is now possible to take out the meaningful/significant patterns from

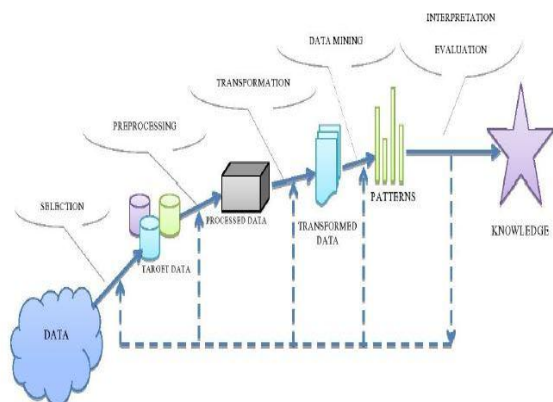


Figure 1 depicts five phases, which are essential in knowledge discovery process.

it. Data mining is of great benefit in a scenario where huge volumes of healthcare data exists.

In scenarios where there is enormous volumes of data that requires categorization, decision-making by individuals is usually poor. Data Mining primarily identifies and extracts the formerly unknown meaningful and key patterns. After that, Patterns' integration follows resulting in knowledge and with the assistance of this knowledge, taking critical impactful and beneficial decisions is possible. Data mining offers a number of benefits. Few of them are as follows: It is very significant in the detection of fraud and abuse, provides improved medical cures at affordable charge, early disease detection, smart healthcare decision support systems etc. They provide improved and enhanced medical services to the patients and assists healthcare firms in taking medical supervision and strategic decisions.

Here are several services provided by data mining techniques in healthcare: approximate number of days a patient will remain admitted in a hospital, hospitals ranking, fraudulent insurance claims by patients as well as by insurance firms, identifies better treatments methods for a specific category of patients, creating effective drug recommendation systems, etc.

Besides these, there are few more aspects which describes the Importance and applicability of Data Mining in Medicine and Public Health such as Data overload, Evidence-based medicine and prevention of hospital errors, Policy-making in public health, Early detection and/or prevention of diseases, Non-invasive diagnosis and decision support, Adverse drug events (ADEs), More value for money and cost savings, Early detection and management of pandemic diseases and public health policy formulation..

III. MINING MEDICAL DATA: ISSUES

Data mining is primarily concerned with describing but not explaining the patterns and trends. But healthcare domain requires those explanations because a minor variation could decide the fate of the patient. It is found that, in majority of the data mining researches on disease and treatment, the conclusions were almost unclear and wary. Few have produced positive outcomes but in addition they have recommend Data Mining in Healthcare for further study. This indecisiveness specifies the dearth of integrity of data mining in these particular aspects of healthcare. Health practitioners are reluctant to deviate from their traditional way of diagnosis so based on existing evidence convincing them is a bigger issue. In addition, Privacy of patient's date and its ethical usage too is a big complication for data mining in healthcare. Data mining results sometimes end-up giving approximate values. For instance, chances of patient again suffering from same disease in future is around 95% but 5% approximation or error is simply unacceptable the

reason being it could turn out to be fatal for patient's life. Finally, Time can be serious constraint in Health care because instantaneous a patient has to undergo treatment and diagnosis.

IV. MINING MEDICAL DATA: CHALLENGES

One of the major hurdle in mining medical data comes from the heterogeneous nature of raw data. This data can be integrated from various diverse sources such as, conversations with patients recorded as data, Lab clinic outcomes and interpretations of these outcomes by specialists. All of these aspects may have an impact on diagnosis and cure of the patient, so they should never be ignored. Before applying any of the DM techniques on the data, one should ensure that some of the important parameters in Dataset are Missing. In other words, incomplete data will not yield proper results in DM. Furthermore, the data in dataset may be incorrect or erroneous also may be inconsistent in nature such data is main hindrance to positive data mining.

In addition to this, It is a tedious task to process Terabytes of records. Information stored is useless if its unavailable in proper format. So various visualization tools are helpful to analyze info in the form of charts and graphs. Therefore, there is a need for medical practitioner to learn and use Visualization tools. There are numerous challenges that must be put forth and addressed before mining on data takes place. Following section briefly describes few challenges for DM in Healthcare Domain: Huge volumes of data: Because of enormous volumes of the healthcare data repositories, existing tools of data mining may need to extract only a section from the data repository. Another method is to handpick few attributes from the data repository In both methods, knowledge about the domain can be used to eradicate unrelated attributes thereby resulting in database size. Inconsistencies in Data: Inconsistencies in data may stem from data entry faults which are usual glitches.

Inconsistencies arrive because of representation of data For instance, if we consider many models for articulating a particular implication, then one model may give different perspective and other may give dissimilar value. In addition to this, the data type may not always echo the exact data type. For instance, a column with numerical data type can represent a minimal variable preset with figures instead of a continuous variable. This plays a crucial role in computations of mean and variance, i.e., statistical analysis.

Improper Integration: Healthcare data is split and spread among clinics, insurance firms and government sectors. This throws a considerable challenge for integration of data in data mining in regard to the confidence that can be put in the outcomes and the semantics of a derived rule. One can use uniform data dictionary and criteria to integrate data from

various dissimilar data sources. Incomplete/Missing data: More often than not, Medical data repositories systems do not gather all the data which is must in KDD process few data attributes goes uncollected due to lapse, inappropriateness, additional risk or wrongness in a particular clinical environment. Few learning approaches like logistic regression, a whole set of data objects may be essential. Even when the techniques take absent values, the data that was uncollected may have autonomous information value and should never be overlooked. One probable method for dealing with missing data is to include alternate values (with most likely values) as missing values.

V. CONCLUSION

Currently, Data mining is still considered to be in its earlier stages. As new diseases continue to emerge, new issues and challenges in Data mining will also pop up. In Health care, especially when it comes to timely sharing of data between parties involved there is always scope for improvement. To be precise, healthcare firms and central bodies has to create or lay a framework for robust standards for data quality in order to provide scope for favorable or encouraging research. Data mining techniques has enormous ability to increase the precision of the medical system results.

REFERENCES

- [1] Anusha N, Rajashree, Srikanth Bhat K "A Survey On Medical Data By Using Data Mining Techniques".
- [2] E. Mercy Beulah, S. Nirmala Sugirtha Rajini, N. Rajkumar "Application Of Data Mining In Healthcare: A Survey",2016
- [3] Hosseinkhah, Fatemeh, Hassan Ashktorab, And Ranjit Veen. "Challenges In Data Mining On Medical Databases." Database Technologies: Concepts, Methodologies, Tools, And Applications. Igi Global, 2009. 1393-1404.
- [4] Țăranu, Ionuț. "Data Mining In Healthcare: Decision Making And Precision." Database Systems Journal 6.4 (2016): 33-40.
- [5] Durairaj, Manoj, And Veera Ranjani. "Data Mining Applications In Healthcare Sector: A Study." International Journal Of Scientific & Technology Research 2.10 (2013): 29-35.
- [6] Ahmad, Parvez, Saqib Qamar, And Syed Qasim Afser Rizvi. "Techniques Of Data Mining In Healthcare: A Review." International Journal Of Computer Applications 120.15 (2015)..
- [7] Dr. C. Sunil Kumar, Dr. A. Govardhan, B. Sunil Srinivas "Data Mining Issues And Challenges In Healthcare Domian, 2014.
- [8] Sujatha, R., R. Sumathy, And A. R. Nithya. "A Survey Of Health Care Prediction Using Data Mining." International Journal Of Innovative Research In Science, Engineering And Technology 5.8 (2016): 14538.
- [9] Sujatha, R., R. Sumathy, And A. R. Nithya. "A Survey Of Health Care Prediction Using Data Mining." International Journal Of Innovative Research In Science, Engineering And Technology 5.8 (2016): 14538.

Authors Profile

Rama Krishna K has around 10.5 years of experience in teaching profession, Industry and research. I am working as Assistant Professor and currently pursuing PhD at Presidency University on Data mining. My objective is to pursue a challenging career in an organization where I can use my skills and experience for the benefit of students, organization and for my professional growth. To actively involved in education & research activities of the university and to explore more on the current research trends in the field of Data Mining.

K G Mohan has in total Experience of 33 years in Teaching and Administrative experience working in different cadre like Assistant Professor, Associative Professor, Professor, Head of the Department, Dean-R&D, Dean-Academics and Principal