

# A Comprehensive Review of Current and Future Applications of Data Mining in Medicine & Healthcare

Dr.Sandeep Kautish<sup>1</sup>, Prof. Rana Khudhair Abbas Ahmed<sup>2</sup>,

<sup>1</sup>Principal - Engineering, North West Institute of Engineering & Technology, DhudikeTakhawadh Link Road,VPO Dhudike, Moga, Punjab, India 142053

<sup>2</sup>Professor, Alrafidain University College/ Computer Engineering Techniques Department, Baghdad, Iraq.

**Abstract**—Data mining is one of the techniques to find hidden information in large amounts of data and it has been widely used in many areas as in communication, credit assessment, marketing, health and medicine, stock market prediction, knowledge acquisition, hazard forecasting, banking, scientific discovery, education, fraud detection, etc., but data mining is significantly applied to medicine for the diagnosis of several diseases such as skin cancer, breast cancer, lung cancer, diabetes, liver disorder, heart disease, kidney failure, kidney stone, hepatitis etc. This paper discusses the data mining applications in medical and healthcare industry including analysis of data for better gaining results in preventing the occurrence of various errors in hospitals, early detection and prevention of various diseases and saving more lives by reducing death rates. The success of medical data mining depends on the provided clean medical data resources.

**Keywords** - Data mining, Medicine, Applications, Healthcare, Current.

## I. Introduction

Data Mining is one of the most motivating areas of research with the objective of finding meaningful information from huge data sets [1, 2]. In present era, Data Mining is becoming popular in healthcare field because there is a need of efficient analytical methodology for detecting unknown and valuable information in health data [2, 3]. Data mining techniques such as classification, clustering and association are used by healthcare organization to increase their capability for making decision regarding patient health [3]. Various studies highlighted that Data Mining techniques help the data holder to analyze and discover unsuspected relationship among their data which in turn helpful for making decision [2, 4].

## II. Data Mining in Medical and Healthcare

Today, health care industry generates large amounts of data about hospital resources, patients; diagnosis of diseases, medical devices, electronic patient records etc., [5], and this large amount of data is to be processed for knowledge extraction that enables support for cost-savings and decision making [5, 6]. Data mining introduces a set of methods that can be

applied to this preprocessed data to discover hidden patterns that provide healthcare professionals an additional source of knowledge for making clinical and administrative decisions [5, 6].

The following describes data mining in various medical applications:

- A. Diagnosis:**Data mining can assist in decision making with a large number of inputs and in stressful situations. It can perform automated analysis of Pathological signals (ECG, EEG and EMG) and Medical images (ultrasound, mammograms, CT, X-ray, and MRI) [5, 6].
- B. Therapy:** Based on modeled historical performance, data mining can select best treatment plans [5, 6].
- C. Prognosis:**Accurate prognosis and risk assessment are essential for improved disease management and outcome [5, 6].
- D. Biochemical/Biological Analysis:**Data mining can automate analytical tasks for urine and blood analysis, tracking the level of glucose, determining level of ion in body fluids, pathological condition detection [5, 6].
- E. Epidemiological Studies:**Study of health, morbidity, injuries disease and mortality in human communities. It can discover patterns relating outcomes to exposures, study independence or correlation between diseases and analyze public health survey data [5, 6].
- F. Hospital Management:**Optimize allocation of resources and assist in future planning for improved services [5, 6].

## III. Developments till now: Current Applications

Table (1) illustrates the current applications of data mining in medicine for the last ten years.

Table (1): current applications of data mining in medicine.

SL No.	Author and year of Research	Explanation of Application	Limitations
1.	Md. Rafiqul Islam et. al., 2005.	In order to learn the attributes of lung cancer, the authors implemented data mining in image archiving systems using an algorithm based on an inductive decision tree. The images were preprocessed and feature subset was selected before applying the algorithm in order to enhance the accuracy of experiment and reduce unnecessary data [6, 7, 8].	Abundance of image data was found, but it can be solved by applying proper data mining techniques which can help to extract the attributes necessary to study any disease with accuracy [6, 9, 8].
2.	AbdelghaniBellaachia et. al, 2006.	In order to predict breast cancer, the authors compared between the results of three techniques, the back-propagated neural network the Naïve Bayes and the (C4.5) decision tree algorithms [6, 8, 9].	A decision tree is similar to flow chart having a tree structure where tests are performed at each level except the last one which gives the result. It is a powerful means for classifying a data set. Predicting the survival time of a cancer patient by this means can actually help the doctor to plan further treatment procedures and the patient to take care of himself so that the time can be extended as much as possible [6, 8, 9].
3.	Ghim-Eng Yap et. al., 2007.	Whenever dealing with data sets errors are evident.	1. If the biological data was noisy then it

		These errors may have effect on the output of results when not diagnosed and this effect may be very serious when it's about the medical diagnosis of a disease. The authors made an attempt that deal with this issue using a knowledge discovery approach by Bayesian network learning [6, 9, 10, 11]. Also a novel procedure for handling the error is implemented to deal with the uncertainty caused by them [8, 10, 11].	will lead to a major problem in the diagnosing result and while managing this noise can give equally good result in cancer detection even with the use of only ten proteins. Such type of tool discussed can prove to be very efficient in screening the patients [6, 9, 10].
4.	Dimitris Bertsimas, 2008.	The authors made an attempt to estimate the medical treatment cost and using the past cost pattern to predict the future cost of patient disease. The diagnostic, procedure and drug related information are present in the claims. Cost bucketing is performed on the samples and the Baseline method is used along with classification trees and clustering to group the members with similar cost characteristics [6, 8, 12].	Care must be taken of the errors for accurate prediction of medical costs. Here the measures of errors used are the hit ratio, the penalty error and the absolute prediction error (APE). This method can be used by the insurance companies too for pricing the health insurances. Also the comparison with past expenses can be beneficial in predicting the rise in future too [6, 8, 12].
5.	Ruban D. Canlas Jr, 2009.	The author detected fraudulent insurance claim through data mining and could analyze better health-policy making health centers	Application of data mining techniques in medical field is a big challenge due to the changing behavior

		[6, 8, 13].	from patient to patient. But still data mining algorithms are used quite successfully to extract better information in case of fraud detection in health policies and as well as in early detection of disease such as heart and pandemics from existing data which is stored in the medical database [6, 8, 13].			have used different parameter like confidence factor, pruning. For classification, anomaly detection and automatically categorize medical images on real mammograms two data mining techniques, association rule mining and neural networks is used [9, 16].	amounts of medical data leads to need for powerful data analysis tools to extract useful knowledge. And healthcare system has huge data available but effective analysis of those using proper tools is the job of data mining techniques. Finding out the hidden relationships and patterns in the data can help predicting a disease very accurately [8, 16].
6.	VeenuMangat, 2011.	Data mining was used in medical records to enhance strategic decisions. Weka version (3.6.0) is software used for the data mining analysis. Data set were converted into Weka data format after that different associative algorithm was executed on data set and their results were comparatively analyzed. There are four algorithms in which comparative analysis is done. After comparative analysis it was found that predictive Apriori is best fitted algorithm in strategic decisions with highest and lowest accuracy values [9, 16, 17].	The authors applied association algorithm to extract association rule, but it is noticed sometimes data sets itself might not be suited for association tasks in data mining. No data mining tools and model will provide 100% accuracy by itself in pure and robust automated system. [10, 17].	8.	Taranath NL et. al., 2013.	The author have used the machine learning approach, automatic learning is used in certain tasks such as medical imaging. In this there are two tasks to be performed: identification and extract informative sentence on disease, fine grained classification according to semantic relation of the sentence on the basis of diseases and treatment. [8, 16].	The goal of paper is to develop prediction model, but researcher has planned to perform additional experiments with more dataset and algorithms to improve the classification accuracy. On the basis of this model that can predict specific heart diseases [8, 20].
7.	ShwetaKharya, 2012.	The author applied decision trees in breast cancer. To generate best decision tree be	The authors have discussed the issue that availability of huge	9.	Matthew Herland et. al., 2014.	The authors discussed the concept of big data analysis for improving the quality of healthcare and also discussed the problems regarding analyzing such huge data in a reliable manner and explored big data tools and approaches at various levels	The frequency of occurrence of diseases varies with geographic area, habits of people, season etc. [8, 19].

		which deal with live molecular, tissue, patient and population [8, 18].	
10.	KasraMadadipouya, 2015.	A new algorithm was implemented based on (C4.5) to mind data for medical applications and then it is evaluated against two datasets in order to improve the classification accuracy and deepness of tree in constructing decision trees, in comparison with (C4.5). During choosing the splitting criteria, the proposed algorithm selects two attributes simultaneously, not only one which enables algorithm to discover the greater information gain ratio of the criterion in the role of the splitting node of decision tree. The result of testing of the proposed algorithm proved that the classification accuracy in the generated decision tree is improved. The proposed algorithm had reduced the depth of the tree [20].	The proposed algorithm has some weaknesses such as more time for computation and it is not still able to escape from being trapped into local optimum [20].

**IV. Developments in Process: Future Applications**

Data mining has a great future and in order to increase the accuracy of the data mining medical system results, we can combine more than one data mining technique together with genetic algorithms, fuzzy logic, artificial neural networks and machine learning and make the system as domain specific, for specific type of diseases such as kidney cancer and making system user friendly to get more feedback. Also data mining can focus on graph; multi

represented objects, noise, privacy concerns and security. Future focus must be on evaluating algorithms on several datasets and can reduce time for computation by increasing the information being trained.

**V. Conclusions and Challenges**

Data mining algorithms have excellent results in medical science and had shown great results for the past ten years. Data mining techniques have been used to make successful decisions that will improve success of healthcare organization and health of the patients. Combining more than one data mining technique for diagnosing or predicting diseases could yield more promising results. There are some techniques that were used in data mining such as artificial intelligence, machine learning, pattern recognition and statistics to serve business, web, and healthcare sectors. Data mining challenges are in data security, accessibility, incompleteness, inconsistent, noise and privacy concerns.

**VI. REFERENCES**

- [1] Thenmozhi K. , P.Deepika, "Heart Disease Prediction Using Classification with Different Decision Tree Techniques", International Journal of Engineering Research and General Science, Vol. 2, Issue 6, October-November 2014, pp. 6-11.
- [2] Sheetal L. Patil, "Survey of Data Mining Techniques in Healthcare", International Research Journal of Innovative Engineering, Volume1, Issue 9, September 2015, pp. 1-3.
- [3] H. C. Koh and G. Tan, "Data Mining Application in Healthcare", Journal of Healthcare Information Management, spring, vol. 19, issue. 2, 2005, pp. 64-72.
- [4] J. Han and M. Kamber, "Data mining: concepts and techniques", 2nd edition. The Morgan Kaufmann Series, 2006.
- [5] Gayathri. P, Dr. N. Jaisankar, "Medical Data Mining Applications and its Uses", International Journal of Engineering Technology Science and Research IJETSR, Volume 2, Issue 6, June 2015, pp. 8 -14.
- [6] NiyaWerts, Monica Adya, "Data Mining in Healthcare: Issues and a Research Agenda", the proceedings of the conference AMCIS 2000 Proceedings, Association for Information Systems (AIS), 2000.pp. 94-97.
- [7] Md. Rafiqul Islam, Morshed U. Chowdhury and SafwanMahmood Khan, "Medical Image Classification Using an Efficient Data Mining Technique", Complexity International, Dhaka, Bangladesh, vol. 12, 2005, pp. 01-09.
- [8] Amrita Kundu, Pallavi, " MEDICAL DATA MINING: A REVIEW ", International Journal of Advanced Technology in Engineering and Science, vol. 03, special issue 01, march 2015, pp. 898-908.
- [9] AbdelghaniBellaachia, ErhanGuven, " Predicting Breast Cancer Survivability Using Data Mining Techniques ", Washington DC 20052, The George Washington University, 2006, pp. 01-04.
- [10] Ghim-Eng Yap, Ah-Hwee Tan, Hwee-Hwa Pang, " Learning Casual Models for Noisy Biological Data Mining: An Application to Ovarian Cancer Detection ", Association for the Advancement of Artificial Intelligence, Singapore, 2007, pp. 354-359.
- [11] Petra Kralj, Nada Lavra, DraganGamberger, AntonijaKrsta, "Contrast Set Mining for Distinguishing Between Similar Diseases ", chapter book: Artificial intelligence in medicine, Springer Berlin Heidelberg, series vol. 4594, 2007, pp. 109-118.