



The Necessity of Data Mining in Clinical Emergency Medicine; A Narrative Review of the Current Literature

Elahe Parva¹, Reza Boostani², Zahra Ghahramani^{3*}, Shahram Paydar³

¹Technical and Vocational University, Shiraz, Iran

²Biomedical Engineering Group, CSE & IT Dept., ECE Faculty, Shiraz University, Shiraz, Iran

³Trauma Research Center, Shahid Rajaee (Emtiaz) Trauma Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

*Corresponding author: Zahra Ghahramani

Address: Trauma Research Center, Shahid Rajaee (Emtiaz) Trauma Hospital, Shiraz University of Medical Sciences, Shiraz, Iran. Tel: +98-71-36360697;

Fax: +98-71-36254206

e-mail: ghahreman2@yahoo.co.nz

Received: January 14, 2017

Revised: February 20, 2017

Accepted: February 24, 2017

ABSTRACT

Clinical databases can be categorized as big data, include large quantities of information about patients and their medical conditions. Analyzing the quantitative and qualitative clinical data in addition with discovering relationships among huge number of samples using data mining techniques could unveil hidden medical knowledge in terms of correlation and association of apparently independent variables. The aim of this research is using predictive algorithm for prediction of trauma patients on admission to hospital to be able to predict the necessary treatment for patients and provided the necessary measures for the trauma patients who are before entering the critical situation. This study provides a review on data mining in clinical medicine. The relevant, recently-published studies of data mining on medical data with a focus on emergency medicine were investigated to tackle pros and cons of such approaches. The results of this study can be used in prediction of trauma patient's status at six hours after admission to hospital.

Keywords: Necessity; Data mining; Clinical; Emergency medicine.

Please cite this paper as:

Parva E, Boostani R, Ghahramani Z, Paydar S. The Necessity of Data Mining in Clinical Emergency Medicine; A Narrative Review of the Current Literature. *Bull Emerg Trauma*. 2017;5(2):90-95.

Introduction

The duty of medical sciences is to treat ill health and promote good health in the community. Most requirements to achieving this goal are knowing the condition of the body, its responses to external and internal stimuli, and how they influence the internal and external factors of body's system. With a deeper understanding of the human body, the actions and interactions of its various organs

will be determined and understood. Nevertheless, these achievements are limited within the interval of possible experiments. Experimental results reveal the reflex of different organs (subsystems) into internal and external inputs within the body. Therefore, as soon as a change becomes stronger or wider, the number of subsystems involved in creating responses will increase. It should be noted that the environmental impacts are not often individual, and the factors that influence

their interaction with subsystem will cause more complex responses to be issued by the body. This matter will be more complicated when the body is changed by underlying diseases in the normal population. This state will cause different responses and difficulties for predicting the body's changes.

What a simple analysis gave to scientists in this field helped them understand and form conclusions based on medical information, was helpful for understanding the basic terms of the system on the whole, and aided in identifying and understanding the factors. With the possibility of collecting a patient's information and the emergence of big data in several areas related to health, statistical tests lost their ability to analyze the situation and identify main factors [1]. As mentioned above, determining the complexity of the body and the interaction of this complex system is often accompanied by multiple external factors. This is followed by an ineffective statistical system in understanding and predicting reliable conditions, especially with multiple time variables and parameters [2]. The need for a top analysis system coincided with the emergence of data mining - the process of knowledge discovery- which was a mixture of machine learning, expert systems, statistics, etc. Such system showed a better understanding of the process and prediction of the future performance of complex systems with analysis of their efficiency in economic and military fields. In particular, data mining can demonstrate the underlying patterns of a system along with the functions of each subsystem in the face of changes [2]. The main goal of data mining is to extract hidden knowledge from a very large sets of data which is not possible to observe them with simple statistical analysis [3]. In fact, the data mining process makes it possible for owners of big data to better understand the dependency among the attributes of the samples in a big dataset and interpret the subsystem processes and to create laws, and predictions of the corresponding subsystem behavior [4].

Data Mining in Emergency Medicine

Emergency medicine is the front line of hospital medical services and is a department that people seek medical care immediately after an emergency. Data mining is a new technique that develop the artificial intelligence and database technique in recent years. It is focusing on database re-analysis including the aim of discovering the valuable information about unknown databases and also to determine the data pattern [1]. Data mining used in medical related research to explore the reduction of patient complaints which arise from insufficient and improper treatments. Therefore, data mining will upgrade the medical quality and also save the waste of medical resources. Shi *et al.*, [5] showed that emergency triage and the scheduling shift of physicians by using data mining analysis will

reduce the classification of noises and determine the classifying levels of triage by classification. Data mining technique will increase the consistence of triage classification in emergency medicine where they used three techniques of data mining to increase this consistence [6]. Computer system can be used to generate calls for reservation. Also, they found that data mining of patient's treatments will help to inform thinking the nature work of emergency departments. The thinking by process-based were used to derive a simple model of emergency department operation [7].

The Ways in Which Data Mining Helps Medical Sciences

In general, the areas of medical sciences that require data mining analysis can be categorized into the following items:

Identifying the complex mechanisms of different body subsystems and their interactions with each other [8, 9];

Identifying people who are at risk for diseases of a genetic predisposition or caused by environmental factors [10];

Identifying disease mechanisms and their interactions with the problems of the body [9];

determining disease prognoses, and facilities management [11];

Establishing decision support systems to make the best decision, especially when the disease is multi-factorial, when more factors are involved in determining the course of the disease, in emergencies, or in acute phases of a disease [10, 11];

Evaluating diagnostic and treatment tasks and relationships and identifying shortcomings and capabilities [12];

Finding the best screening methods for diseases and injuries, particularly for patients in critical conditions [13].

Data mining is the result of using implemented algorithms in software to cover the needs of medical science in each section with the construction of analytical models, categorizing, information prognosis (prediction), and presentation. There are different techniques in data mining, but the following subjects will be used more in the discussion of analytical or predictive medicine: Classification, Regression, Clustering, Discovery the interpretable rules of dependencies, and Sequences Analysis (Table 1) [14].

The History of Data Mining in Medical Sciences

Nearly 10 years after the emergence of a data mining process in the field of trade, communication management and analysis of crime, it was first deployed in the field of health in the early 1990s. Apparently, data mining was used to identify trends, income and expenses of treatments; data

Table 1. Categories of data mining methods and algorithms used in Medical Sciences

No.	Medical Field Research	Data mining method	Data Mining Algorithm	Description
1	Disease Prediction	Categories Clustering Rules of dependence	K-means Apriori	Determine the factors affecting cancer types
2	Determine the best type of treatment	Categories Clustering Rules of dependence	K-means Apriori	Determine the best type of cancer treatment through 3 methods: Surgery, Chemotherapy, Radiotherapy
3	Experimental Data Evaluation	Clustering	SAMBA	The study of genes behavior with DNA strands to predict genetic disorders and fatal diseases
4	Prediction in Emergency Patients	Clustering	K-mean combination and Neural Network SOM ^a Algorithm	To take proper and timely treatment decisions and reduce hospital costs
5	Length of Stay Prognosis	Categories Clustering	Decision Tree Neural Network Naïve Bays C 4.5	Anticipated duration of hospitalization digestive patients who need short-term care to reduce hospital costs
6	Identify and predict of disease symptoms	Clustering	Combination of Genetic Algorithm and K-means Algorithm	Identify and prediction of heart attacks
7	Traditional Chinese Medicine	Clustering	SVM ^b Decision Tree Bayesian Network	Discover the different syndromes
8	Traditional Chinese Medicine	Rules of dependence	CNA ^c	Find good points of Acupuncture and patterns of medicinal plants
9	Traditional Chinese Medicine	Clustering	SVM	Prediction of starting diabetic neuropathy
10	Traditional Chinese Medicine	Clustering	Neural Network	Treatment of rheumatoid arthritis

^aSOM: Self-organizing Map; ^bSVM: Support Vector Machine; ^cCNA: Complex Network Analysis

mining capabilities were studied for monitoring and understanding clinical data [7, 15].

Along with the progress of the data mining process in computer science and statistics in medical sciences, evidence-based medicine was raised which was noticed by a medical team to generate usable data and create lots of information. Therefore, knowledge was supposed to be extracted from this information. At this time and because of the inability of linear statistics to produce information knowledge, data mining was considered as a tool for knowledge discovery [16-18].

In the years between 1997 and 1999, Prather [17] and Babic [18] showed the importance of data mining in monitoring medical information with an emphasis on a large volume of data. Almost from the time of publication of these articles and the clarification of the role of data mining in investigating medical information, a two-way communication was formed between specialists in the field of data mining and those who involved in the health sector. These people found the ability to explore relationships among data and predict complex processes. Data mining experts had found the area where there was big data, and data was studied on a regular basis under different conditions and with different parameters and variables. Their so far joint cooperation has led to impressive and operational achievements [19, 20].

Over time, data mining has almost found its position in medical sciences. Bellazzi released a guideline for data mining in medical sciences with the aim of changing attitudes about the task of predicting

patients' conditions in various fields [21]. In 2003, Hripesak [22] showed that the use of data mining can increase patient safety and prevent medical errors. Lynch [15] states the importance of data mining and its ability to solve medical problems according to the capabilities of these techniques.

Discussion

In recent years, several studies have been carried out using data mining schemes in different medical fields. Engineers have evaluated the adequacy of data mining algorithms and models in different areas of health. Based on the different aspects mentioned at the beginning of this article, some studies are discussed below:

A lot of data mining research project has been made on identifying the complex processes of the body, especially at the molecular level. This matter was originally considered by experts, especially with the advent of new technologies allowing people to have access to genetic information. Researchers have gathered a great deal of information about different gene sequences which can be analyzed with data mining techniques, and new knowledge in the field of system performance can be achieved according to their genetic formulation [16].

In the year 2012, Tilton successfully provided a software using genetic data through data mining which could model and predict the morphology of conceiving its genetic formula through analyzing data related to mRNA. This software is currently

used and is being upgraded [23].

Regarding the identification of people at risk, diseases such as cancer form one of the most popular data mining application areas in medicine from the aspects of genetic predisposition and environmental factors. Data mining is considered in the first days because of the possibility of analyzing various aspects of the situation of people with cancer and the importance of identifying the conditions that endanger people. In the early years, most clinical and radiological information of a patient was assessed using the cognitive development index and genetic information; data mining is mostly used to monitor the massive amount of data obtained from genetic analyses. In the year 2001, Kuo and Chang reviewed and categorized the results of sonographic findings in patients with breast cancer based on a decision tree [24]. They were able to find patients with breast cancer through data analysis and invented a forecasting system upon this tree and the findings of ultrasound examination and imaging precision [25]. Then, they designed a software system to predict malignant breast masses with ultrasound findings [24, 25]. Asadi *et al.* [10] studied the factors leading to the development of cancer using data mining techniques and detected the associations between these factors with the information recorded in the cancer registry of Nemazee hospital in Shiraz city, Iran.

The latest achievements of data mining in the field of cancer focus more on the data of genomes and genetic proteins such as RNA and DNA.

In the year 2015, Moore successfully achieved patterns using a model which could predict a person's risk of cancer with information from the mRNA analysis. He was also able to find the initial prediction model using the data mining process [16].

In 2016, Milioli successfully classified breast cancer using data mining through a genetic data bank. This work resulted in a better understanding of the pathology and diagnosis of cancer [26].

Mechanisms of disease and how the body interacts with the problem have been the focus of several attempts. Dehghani *et al.* [27] used clustering techniques to identify patients with heart attack and to predict their heart attack by the combination of K-means and genetic algorithms. In another study, Apitus successfully predicted the incidence of neurodegenerative diseases through a mathematical model optimized by genetic algorithm from non-affected persons [28]. In the year 2016, Dipnall [29] successfully determined much significant data related to experimental data in patients with depression using data mining. A typical analysis of its relationship had not been discovered. Dipnall was able to find a significant relationship between the laboratory findings and the risk of depression through the use of a hybrid system with a mix of linear analysis and data mining techniques. Paydar *et al.*, [30] by modeling could predict malignancy in thyroid nodules.

Determining diseases and facilities management, including operational applications of data mining in medical fields, is another topic, the importance of which will be raised, especially in the limited time of care facilities, high cost of care, or low probability of response to the treatment. In 2011, Lin *et al.* [11] defined the conditions and outcomes that would lead to death despite the cost. The study examined the data of patients admitted to the emergency room, and follow-up costs through clustering patients and use of the nervous system.

The study of Delen *et al.*, [12] modeled the prognosis of patients after lung transplantation through machine learning by using demographic data, operating conditions, and paraclinical findings before, during, and after transplantation. The purpose of modeling was to predict postoperative lung transplant candidates; through modeling, they were able to predict the likelihood of death in lung transplant patients. Therefore, subsequent decisions to perform very heavy and costly lung transplants could be made.

The progress of using data mining in data analysis and the interpretation of genetic explanations of humans caused challenges to predicting the risk of people for a disease, especially in cases where there was little predictive power or methods were not selected by appropriate categories. In 2006, Moore's study found the needs to enhance the accuracy of the models and to ensure a high degree of confidence [12, 31].

The existence of a decision support system to make better business decisions is important for researchers and decision-makers in the field of health. With the use of a decision support system, it is likely that fewer errors will occur, better decisions will be made, and better results will be achieved. In 2009, Canlas unveiled the efforts of Thungarel and Gorunescu who used the K-means clustering algorithm for diagnosing and treating cervical cancer [32].

The breast cancer diagnosis decision support systems proposed in 2001 by Kou *et al.*, [25] have been previously discussed. Kou *et al.* have also done much to advance the use of decision support systems in emergencies, including decision support systems that McKenna and Chen [33] used in 2008 to quickly identify trauma patients suffering from blood loss or shock. These researchers provided a system that can accurately identify patients at risk for shock by using an ensemble classifier. They were able to offer an acceptable model using data mining [26, 34] because of the inadequacy of regular monitoring and the effective early detection of patients who were at risk for shock [35-37].

From another angle, there still exist some limitations regarding data mining techniques which prevent us to catch wonderful results in several cases. For instance, DB-SCAN is a very powerful clustering algorithm and is applicable for big data analysis like medical databases but it is not applicable for high

dimensional data like bioinformatics data. On the other hand, K-Means is a good clustering algorithm but its performance is highly dependent on its distance function. Support vector machine (SVM) and neural network have amazing classification capabilities but each of them has its own shortcomings. For instance, SVM provides a statistical generalization capability and does not have any problem to handle high dimensional samples but its performance fails to handle more than 1000 samples. The main weakness of neural networks is that they over-fit on the training data and their test and train results have a big difference in some cases. Fuzzy classifiers and clustering techniques can work well in the uncertain environment, providing interpretable rules and can work well even with insufficient number of samples but they do not have any statistical generalization support and their train and test results might have a drastic difference in some applications. The same story repeats for regression methods in a way

that adaptive boosting regression (Adaboost.R) and support vector regression (SVR) are great to encounter with low number of samples but they fail to handle big data. On the other hand, polynomial regression methods optimized by least square criterion act well in some applications but they do not consider a margin around its regression curve.

In conclusion, we have to confess that there is no perfect method that could handle wide variety of data sets with different specifications, i.e. containing large number of samples, high dimensional samples, data include high portion of noisy samples, two-class and multi-class situations, data with uni-modal and multi-modal distribution, etc. In contrast, there is a growing interest to improve the existing methods by combining and fusing them together to take the benefit from the positive points of each other.

Conflict of Interest: None declared.

References

- Hand DJ, Blunt G, Kelly MG, Adams NM. Data mining for fun and profit. *Statistical Science*. 2000;**15**(2):111-31.
- Theodoraki E-M, Katsaragakis S, Koukouvinos C, Parpoula C. Innovative data mining approaches for outcome prediction of trauma patients. *Journal of Biomedical Science and Engineering*. 2010;**3**(08):791.
- Hassanzadeh M, Razavi Ebrahimi SA. Data Mining Algorithms for Medical Sciences. *Iranian Journal of Medical Informatics*. 2013;**2**(2).
- Moghadasi H, Hosseini A, Asadi F, Jahanbakhsh M. Data mining and its application in health. *Health Information Management*. 2012;**9**(2):297-304. [In Persian]
- Shi YS. Using data mining techniques to analyze and improve for emergency triage and operation of doctor schedule. Master dissertation, National Chin-Yi University of Technology; 2008.
- Lai CH. Data mining applied to the prediction model of triage system in emergency department; a case of medical center in Taiwan. Master dissertation, National Chin-Yi University of Technology; 2007.
- Ceglowski A, Churilov L, Wassertheil J. Knowledge discovery through mining emergency department data. System sciences, 2005. HICSS'05. Proceedings of the 38th annual Hawaii International Conference on. 2005:142c-142c.
- Pardalos PM, Tomaino V, Xanthopoulos P. Optimization and data mining in medicine. *Top*. 2009;**17**(2):215-36.
- Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artif Intell Med*. 2010;**48**(2-3):139-52.
- Asadi N, Sadrodini M. Employing data mining to identify cancer risk factors and determine the optimal treatment in Namazi hospital cancer database. *the*. 2010;**16**:17-8.
- Lin W, Wu Y, Zheng J, Chen M. Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital. *Expert Systems with Applications*. 2011;**38**(9):11078-84.
- Delen D, Oztekin A, Kong ZJ. A machine learning-based approach to prognostic analysis of thoracic transplantations. *Artif Intell Med*. 2010;**49**(1):33-42.
- Craig JB, Culley JM, Tavakoli AS, Svendsen ER. Gleaning data from disaster: a hospital-based data mining method to study all-hazard triage after a chemical disaster. *Am J Disaster Med*. 2013;**8**(2):97-111.
- Hassanzadeh M, Razavi Ebrahimi SA. Comparison of data mining algorithms classification in Medical Sciences. *Iranian Journal of Medical Informatics*. 2012. [in Persian]
- Lynch SM, Moore JH. A call for biological data mining approaches in epidemiology. *Bio Data Min*. 2016;**9**:1.
- Moore AC, Winkler JS, Tseng TT. Bioinformatics Resources for MicroRNA Discovery. *Biomark Insights*. 2015;**10**(Suppl 4):53-8.
- Prather JC, Lobach DF, Goodwin LK, Hales JW, Hage ML, Hammond WE, editors. Medical data mining: knowledge discovery in a clinical data warehouse. Proceedings of the AMIA annual fall symposium; 1997: American Medical Informatics Association.
- Babic A. Knowledge discovery for advanced clinical data management and analysis. *Stud Health Technol Inform*. 1999;**68**:409-13..
- Liao S-H, Chu P-H, Hsiao P-Y. Data mining techniques and applications—A decade review from 2000 to 2011. *Expert Systems with Applications*. 2012;**39**(12):11303-11.
- Silwattananusarn T, Tuamsuk K. Data mining and its applications for knowledge management: a literature review from 2007 to 2012. arXiv preprint arXiv:1210.2872. 2012.
- Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform*. 2008;**77**(2):81-97.
- Hripcsak G, Bakken S, Stetson PD, Patel VL. Mining complex clinical data for patient safety research: a framework for event discovery. *J Biomed Inform*. 2003;**36**(1-2):120-30.
- Tilton SC, Tal TL, Scroggins SM, Franzosa JA, Peterson ES, Tanguay RL, et al. Bioinformatics Resource Manager v2.3: an integrated software environment for systems biology with microRNA and cross-species analysis tools. *BMC Bioinformatics*. 2012;**13**:311.
- Kuo WJ, Chang RF, Chen DR, Lee CC. Data mining with decision trees

- for diagnosis of breast tumor in medical ultrasonic images. *Breast Cancer Res Treat.* 2001;66(1):51-7.
25. Kuo WJ, Chang RF, Moon WK, Lee CC, Chen DR. Computer-aided diagnosis of breast tumors with different US systems. *Acad Radiol.* 2002;9(7):793-9.
 26. Milioli HH, Vimieiro R, Tishchenko I, Riveros C, Berretta R, Moscato P. Iteratively refining breast cancer intrinsic subtypes in the METABRIC dataset. *BioData Min.* 2016;9:2.
 27. Dehghani T, Saleh MA, ali Khalilzadeh M, editors. A genetic K-means clustering algorithm for heart disease data. 5th Conference of Data Mining of Iran, Amirkabir University; 2011.
 28. Hofmann-Apitius M, Ball G, Gebel S, Bagewadi S, de Bono B, Schneider R, et al. Bioinformatics Mining and Modeling Methods for the Identification of Disease Mechanisms in Neurodegenerative Disorders. *Int J Mol Sci.* 2015;16(12):29179-206.
 29. Dipnall JF, Pasco JA, Berk M, Williams LJ, Dodd S, Jacka FN, et al. Fusing Data Mining, Machine Learning and Traditional Statistics to Detect Biomarkers Associated with Depression. *PLoS One.* 2016;11(2):e0148195.
 30. Paydar S, Pourahmad S, Azad M, Bolandparvaz S, Taheri R, Ghahramani Z, et al. The Evolution of a Malignancy Risk Prediction Model for Thyroid Nodules Using the Artificial Neural Network. *Middle East Journal of Cancer.* 2015;7(1):47-52.
 31. oore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol.* 2006;241(2):252-61.
 32. Canlas R. Data mining in healthcare: Current Applications and Issues. School of Information Systems & Management, Carnegie Mellon University, Australia. 2009.
 33. Chen L, McKenna TM, Reisner AT, Gribok A, Reifman J. Decision tool for the early diagnosis of trauma patient hypovolemia. *J Biomed Inform.* 2008;41(3):469-78.
 34. Chen S-S, Haskins WE, Ottens AK, Hayes RL, Denslow N, Wang KK. Bioinformatics for traumatic brain injury: Proteomic data mining. *Data Mining in Biomedicine*: Springer; 2007. p. 363-87.
 35. Lovett PB, Buchwald JM, Sturmman K, Bijur P. The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Ann Emerg Med.* 2005;45(1):68-76.
 36. Friesdorf W, Konichezky S, Gross-Alltag F, Fattroth A, Schwilk B. Data quality of bedside monitoring in an intensive care unit. *Int J Clin Monit Comput.* 1994;11(2):123-8.
 37. Appel R, Bairoch A, Hochstrasser D. 2-D databases on the World Wide Web in methods in molecular biology. *AJ Link*, editor.383-91.