Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100443

Validation of Data Mining Advanced Technology in Clinical Medicine

Abd Elrazek M Aly Abd Elrazek[1,*] *and Ahmed Elbanna*[2]

[1] Department of Hepatology & Gastroenterology, Aswan school of Medicine, Aswan University, Aswan, Egypt.
[2] Department of Computer Science and Technology TU Chemnitz, Germany.

Received: 3 Apr. 2016, Revised: 14 Jun. 2016, Accepted: 27 Jun. 2016
Published online: 1 Jul. 2016

Abstract: The objective of this review study is to assess the feasibility of data mining advanced statistical technology in applied medicine in a way each clinician and IT engineer can understand each other. We have already experienced many studies using data mining with or without traditional SPSS programs in Hepatology Gastroenterology and Endoscopy. A descriptive models were generated using decision tree algorithms (Rapid I, Rapid Miner, Berlin, Germany), accuracies obtained of all studies was significantly high and applicable; ? 90 %. Here we will shed light on the importance of data mining in clinical medicine, analyzing data by advanced data mining technology would discover the significant important leading factors in the applied medicine. Thus the major challenge of biomedical data mining over the next 5-10 years is to make these systems useful to biomedical researchers.

Keywords: data mining, clinical medicine, validation, engineer, clinician.

1 Introduction

Data mining analysis is the process of examining large amounts of data by means of machine learning to create an applicable algorithms. Data mining (DM) as a discipline is largely transparent to the world. Most of the time, we never even notice that its happening. Involuntary or usual actions meaning worth explanation in our life, however we never notice up. For example whenever we sign up for a grocery store shopping card, place a purchase using a credit card, or surf the Web, we are creating every day data. Lying within those data sets are patterns indicators of our interests, habits, time of purchase and even our behaviors. DM allows companies to locate and interpret those patterns, helping them make better informed decisions and better serve their customers. Furthermore theses data are stored in large sets on powerful computers; (super computers) owned by the companies. As we deal with every day, they will be able to understand our mind for purchase, can evaluate each economy. The same go with scientists in technology, medicine, chemistry, astronomy, biology, trading and communication sciences, DM will influence the understanding of science because the major challenge of biomedical research activities is to make these systems useful and applicable to biomedical researchers. The use of DM in applied technology is important to predict factors leading to much innovation and heavy creation by extracting hidden factors never been watched or minded before. In simple words, DM is the process of automatically extracting hidden useful information (interesting patterns) from large databases, thus safe time, money, effort and give much scientific innovations [1,2,3,4,5].

2 An Example of Medical Records in a Hospital

Even medical records increasing heavily as shown below in both vertical and horizontal information (figure 1), accordingly we need such DM to understand correlation, miscorrelation, advantages, disadvantages and important or less important criteria- for each worker and as a team, leading to work promotion, innovation and financial success.

How to apply?

–Large amount of data.
–Information hidden in the data that is not easily evident to humans.
–High computational power.

* Corresponding author e-mail: ahmadrazek@gmail.com

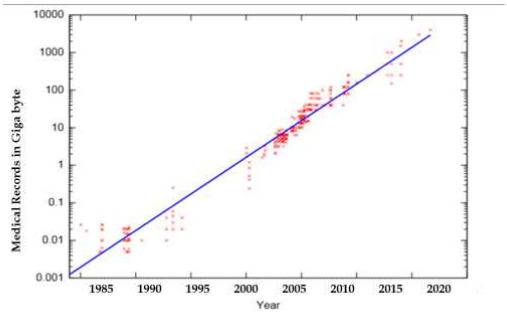


Figure 1

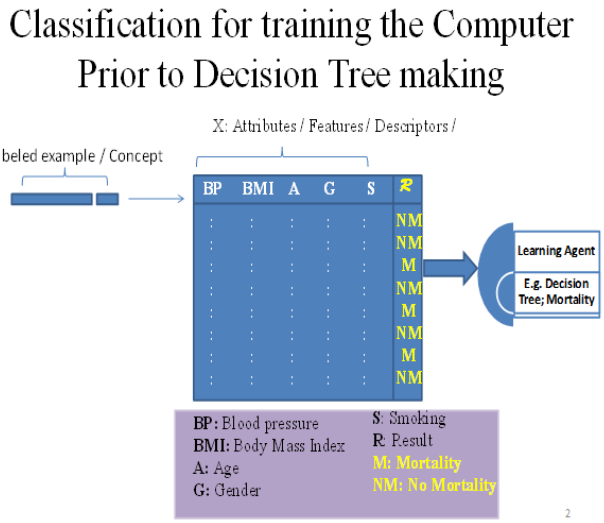


Figure 2

–Give us the worth of the wisdom behind all these data in all scientific fields.

3 Discussion and Review of Previous Work

A research entitled; Mortality prediction of nonalcoholic patients presenting with upper gastrointestinal bleeding using data mining, published in European Journal of Gastroenterology and Hepatology. 2014 Feb; 26(2):187-91. Abd Elrazek et al [6], yielded the most significant mortality predictors; whose patient should be operated immediately (mortality group) and who should be treated medically (non- mortality group), hence we can even predict the life expectancy. The overall accuracy was 92.08%. Other researches using DM we have presented in literature including multi-center studies, are; Detection of

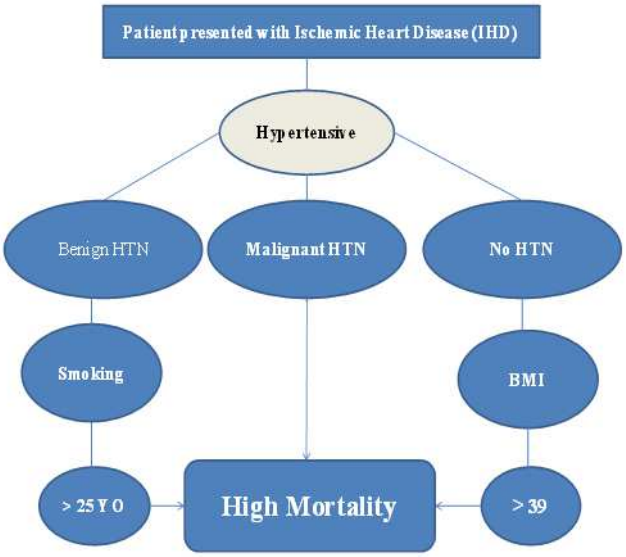
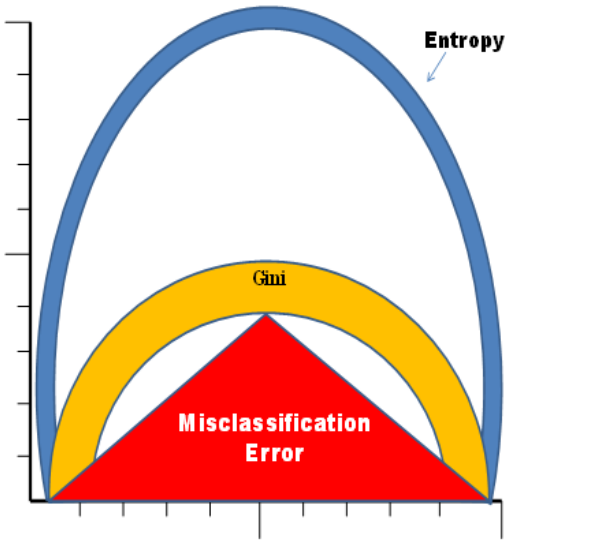


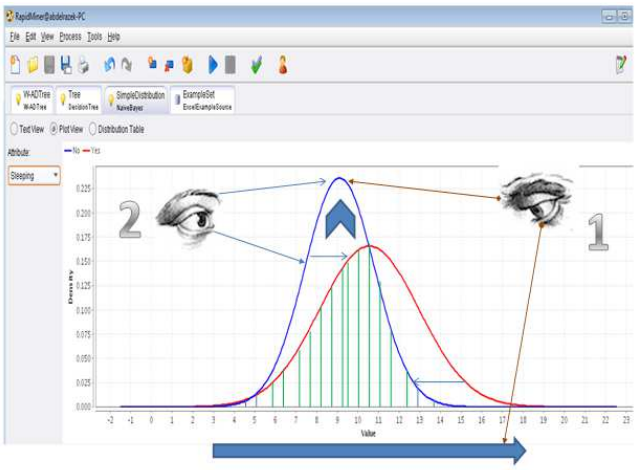
Figure 3



2-Class problem of Comparison Splitting Criteria.

Figure 4

risky esophageal varices using 2D U/S when to perform Endoscopy; American Journal of the Medical sciences [7]. Screening Esophagus during routine ultrasound; medical and cost benefits; European Journal of gastroenterology and Hepatology, [8], Prior to the oral therapy what do we know about HCV-4 in Egypt, a randomized study using data mining computational



Eye of Engineer ;widespread and analytic (1).
Eye of Clinician; focusing and applicable (2).

Figure 5

analysis and the value of U/S to determine priority for EGD in emergency room, Medicine, Baltimore [9,10] and recently accepted manuscript ; Prediction of HCV vertical transmission , what are factors determined using data mining computational analysis; Liver International Journal ; [11] an official Journal of International association for the study of liver (IASL) and others. The use of DM in applied Gastroenterology, Hepatology and Endoscopy was very useful to predict factors leading to both mortality and morbidity by extracting hidden factors never been watched or minded in medicine, DM automatically extracted hidden useful information from large databases.

Understanding Data mining simply; What A clinician Has to understand, What an Engineer has to know?

There are many applications of Data mining including:

–Decision Tree based Methods
–Rule-based Methods
–Memory based reasoning
–Neural Networks
–Naive Bayes and Bayesian Belief Networks.
–Supporting Vector Machines application.

However, the most applicable method in medicine is the decision Tree (DT). DT is a predictive model used to determine which attributes of a given data set are the strongest applicable indicators of a given outcome of a certain disease or functional disorder, furthermore help to categorize data according to importance, also it helps to classify and predict according to data in pattern of a certain disease or a disorder, for example if we want to predict a mortality in a certain group of patients presented with Ischemic Heart Disease (IHD), we have to follow patients passed away due to IHD, Accordingly, the DT Algorithm will be modulated based on mortality (training label decision) in whom expected to get medical complications leading to death, further more different varieties of data influence the DT outcome greatly. When applied these certain data for certain group of patients presented with IHD; training model, the DT showed unexpected predictors-related Mortality figure (2). Additionally shed light the most important predictors related mortality in patients presenting with IHD; surprisingly there were no relation to gender or sex related mortality as cardiologists thought before, however DT learned how to prevent mortality through treatment of hypertensive patients and to decrease their BMI, finally DM guide cardiologists the most important and dependent factors led to mortality, figure (3). Both IT engineer and clinician should be sure that they designed the first row of each the data sets as the attribute of names and save them in the repository with descriptive names, be able to tell what they are. The training label should be written correctly; decision making e.g.: if decision making-related disease mortality in columns 6, it should be written as 6 in label -column repository set. It is important to include all data even if clinician thought useless data. However decision trees have several classifications; Accuracy, Gain-ratio, Information gain and Gini Index. Gini Index Entropy, Classification error and naive distribution computer analysis, shall replace certain mathematical methods, give more accurate and applicable results. Figure (4). Engineer vision is widespread mathematically, however physician vision is more confined. The information obtained from DT should be studied efficiently by both clinician and engineer in order to understand how efficacy those data influence the disease progression or regression in a clinical medical view and how those data input, output and modulation can be changed in a computerized art giving the best results by an intelligence algorithm. Additionally, the engineer has to try varieties of cross-validation aiming to checking for the like hood of false positive in predictive models in Rapid Miner, no theories can be accepted in patients related mortalities until these theories proved clinically (figure 5).

4 Conclusion

We believe that data mining will influence the global trade negotiations, national labor relations and even a breakthrough in evolutionary biology, applicable clinical and academic medicine. DM can predict mortality and morbidity in clinical medicine.

Acknowledgement

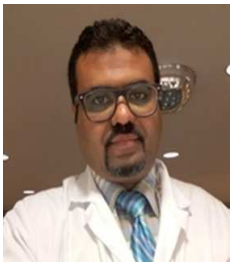
Authors do thank Professor Mansour Kabbash; President of Aswan university for his scientific support and promotion of basic/clinical researches.

Limitation of the study

To simplify the study we did not went deepest in technology science of data mining software programming, in our point of view we presented the most applicable and useful patterns used in different clinical medical fields.

References

[1] Matthew North. Data mining for the mases, 2012.
[2] Abd Elrazek M Ali, H Mahfouz. Prediction of esophageal variceal degrees using data mining: is validated in clinical medicine? Global J Comp Sci Tech 2013; 13:10.
[3] Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008; 77 (2):8197.
[4] Grossi E, Buscema M. Introduction to artificial neural networks. Eur J Gastroenterol Hepatol 2007; 19 (12):1046-1054.
[5] Lu S, Jin B, Cowart LA, Lu X. From Data towards Knowledge: Revealing the Architecture of Signaling Systems by Unifying Knowledge Mining and Data Mining of Systematic Perturbation Data. Plos One. 2013 Apr 23; 8(4):e61134.
[6] Abd Elrazek AE, Mahfouz HM, Metwally AM, ElShamy AM. Mortality prediction of nonalcoholic patients presenting with upper gastrointestinal bleeding using data mining. Eur J Gastroenterol Hepatol 2014; 26 (2):187191.
[7] Abd El Razek M Ali, Mahfouz H, Afify M, fathy Abd wahab et al., Detection of Risky Esophageal varices by 2D U/S: when to perform Endoscopy. Am J Med Sci. 2014 Jan; 347(1):28-33.
[8] Abd Elrazek AE1, Eid KA, El-Sherif AE, Abd El Al UM Screening Esophagus during routine U/S; medical and cost benefits. , Eur J Gastroenterol Hepatol. 2015 Jan; 27(1):8-12.
[9] Abd Elrazek AE, Bilasy SE, Elbanna AE, Elsherif Asim. Prior to the Oral Therapy, What Do We Know About HCV-4 in Egypt: A Randomized Survey of Prevalence and Risks Using Data Mining Computed Analysis. Medicine (Baltimore) 93(28):e204
[10] Ali Hussein AE1, Mahfouz H, Elazeem KA, Fakhry M, Elrazek EA, Foad M, The Value of U/S to Determine Priority for Upper Gastrointestinal Endoscopy in Emergency Room. Medicine (Baltimore). 2015 Dec;94(49):e2241.
[11] Elrazek AE, Omer M, Hawary B, Akhays A, Taher S, Saab S. Prediction of HCV vertical transmission, what are factors determined using data mining computational analysis. Accepted; Liver Int. April, 2016.



Abd Elrazek Mohammad Ali Consultant Hepatologist, Aswan School of Medicine, Aswan University, Egypt Senior Researcher, Clinical and Applicable Data mining, Aswan University, Egypt PHD. Transplant Hepatology, KIFMC, Japan, 2012 PHD. Tropical Medicine, Egypt, 2014 Post Doctorate Fellow, Liver surgery and Transplant Hepatology, UCLA, USA, 2015 Academic Editor; Medicine - Baltimore, USA.

Ahmed elbanna was born in Saudi Arabia, in Medina in 1988. He obtained the Bachelor of Electronics and Communications Engineering from the Arab Academy for Science and Technology and Maritime Transport in Cairo, Egypt in 2011. After graduation he has worked as a technical support engineer in Eitesalat UAE for a year. Later, he joined Vodafone Egypt as a service engineer in the IT operations sector. In 2013, he started his masters degree of Information and Advanced Communication Systems in the Technical University of Chemnitz in Germany. In august 2015 he joined the International Telecommunications Union as a researcher. Currently he is pursuing his masters thesis project in Politechnico di Milano, Italy. ”