PROJECT NAME

Data Mining in Clinical Emergency Medicine

Student name : Farkad Adnan Jaffar – فرقد عدنان جعفر

Faculty: Computer Engineering - Artificial Intelligence and Robotics MSc

1. 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.

Key Words: Necessity, Data mining, Clinical, Emergency medicine

2. 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.

3. Analyzing The System

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].

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 [14].

5. Machine learning and data mining

The dataset represents 10 years (2000-2021) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria.The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medication, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

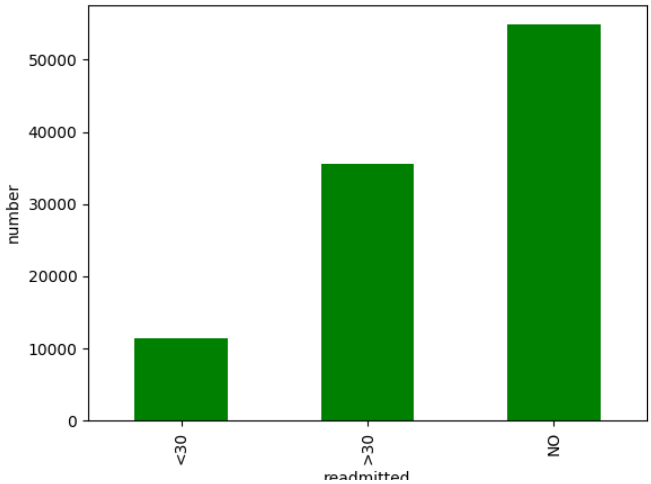
(1) It is an inpatient encounter (a hospital admission).

(2) It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis.

(3) The length of stay was at least 1 day and at most 14 days.

(4) Laboratory tests were performed during the encounter.

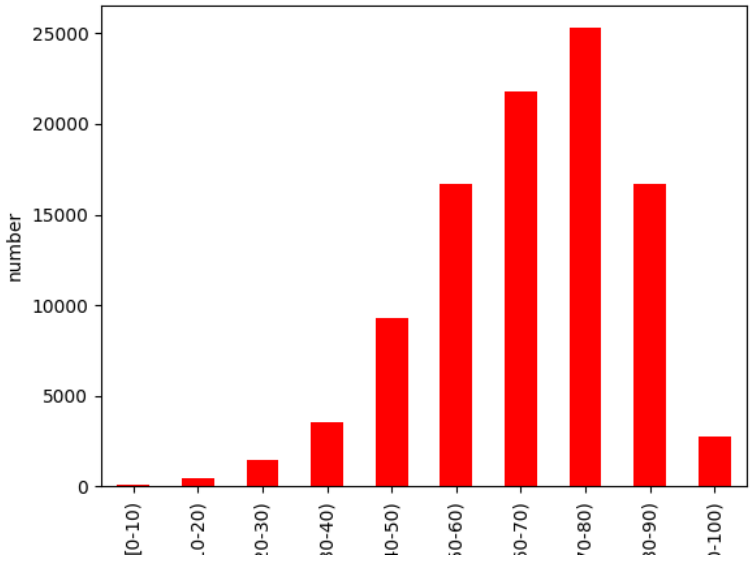
(5) Medications were administered during the encounter.



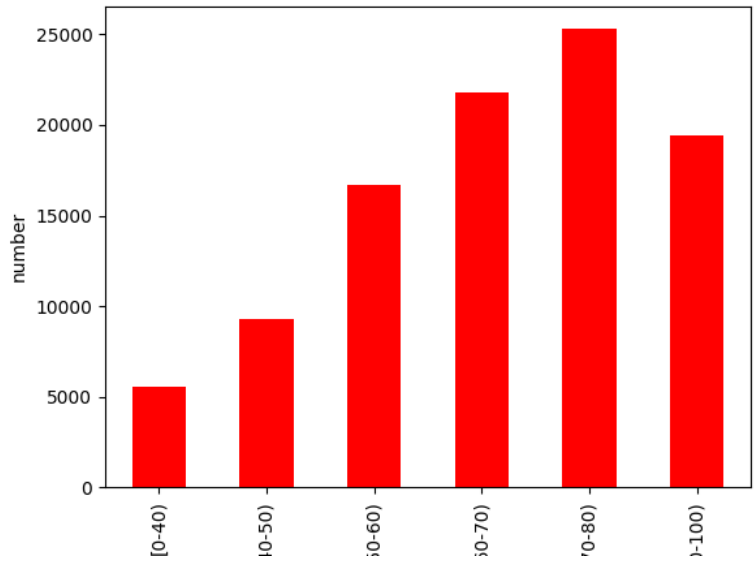
Figure(1) Look at the label:eadmitted,we visualize it by groups

there are three types of labels. We just regroup it to two labels.

data['readmitted'] = pd.Series([0 if val == 'NO' else 1 for val in data['readmitted']])

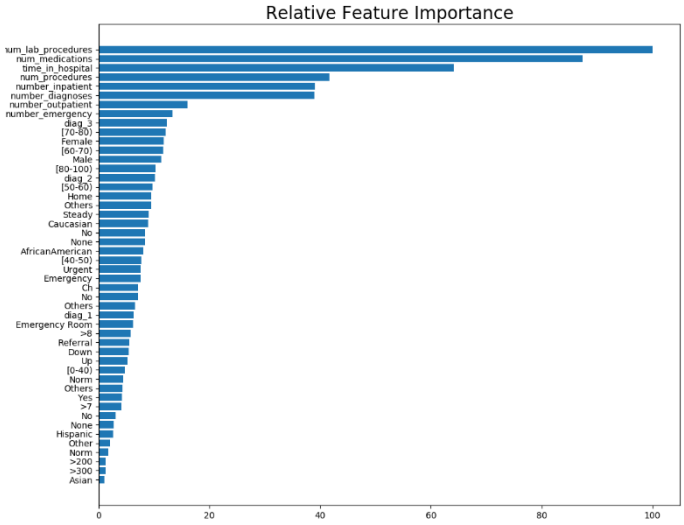


Figure(2) Then we look at the feature:age, we visualize it by groups



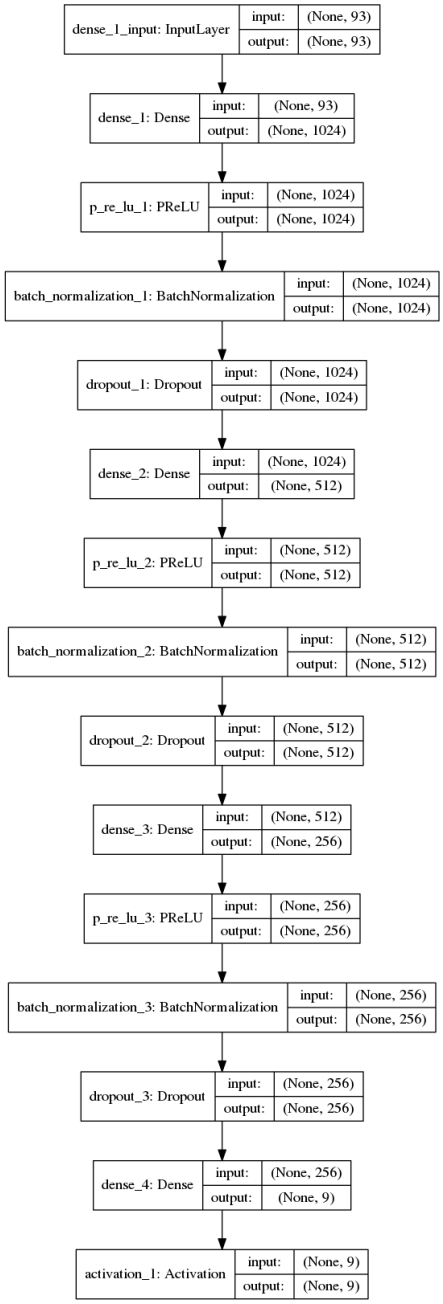
Figure(3) We find it really unbalance,then regroup this feature

There are two many features in this dataset, but many of features are noisy. We use random forest to evaluate the importance of each features and rank them.



A consistent analysis of the performance of our products is crucial. However, due to our diverse global infrastructure, many identical products get classified differently. Therefore, the quality of our product analysis depends heavily on the ability to accurately cluster similar products. The better the classification, the more insights we can generate about our product range.

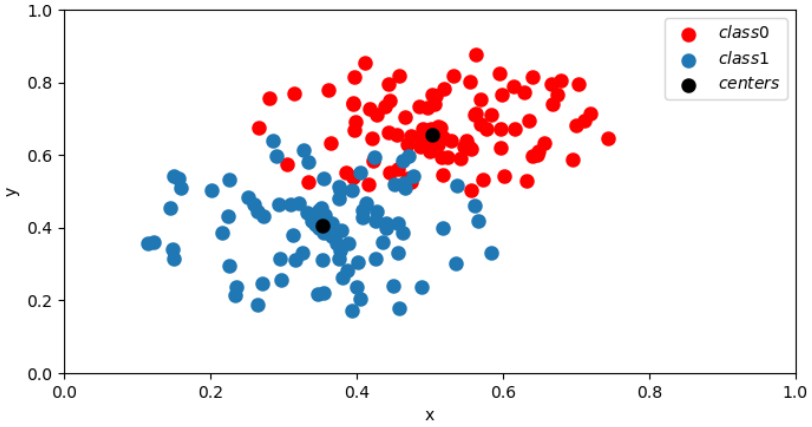
We are going to build a simple ANN to do this work. The structure is shown below



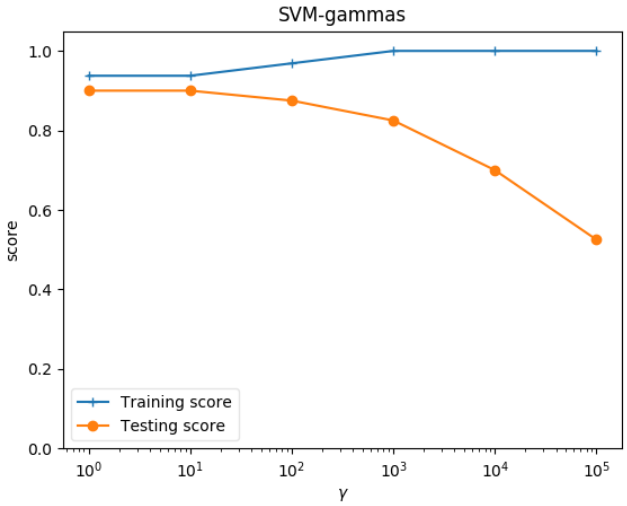
6. The third practice is using SVM

The SVM is an important classify in machine learning, we need to practice how to use the SVC in sklearn lib. You can see all codes in SVM.py

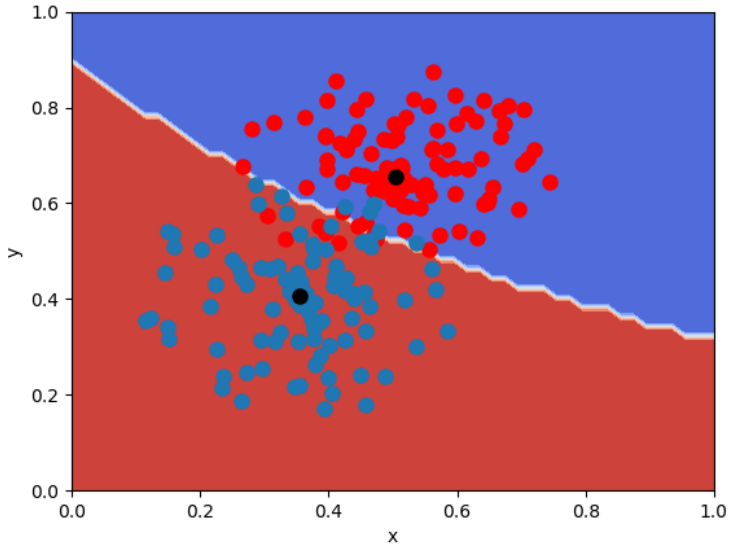
First we create some fake data for classification



Then we use 10 fold cross validation to choose paramater gamma



We can adopt SVC and draw the hyperplane



7. 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].

- # References

[1] Hand DJ, Blunt G, Kelly MG, Adams NM. Data mining for fun and profit. *Statistical Science.* 2020;15(2):111–31.

[2] 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.* 2020;3(08):791.

[3] Hassanzadeh M, Razavi Ebrahimi SA. Data Mining Algorithms for Medical Sciences. *Iranian Journal of Medical Informatics.* 20.20

[4] Moghadasi H, Hosseini A, Asadi F, Jahanbakhsh M. Data mining and its application in health. *Health Information Management.* 2020;

[5] 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; 2020

[6] 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; 2020.

[7] Ceglowski A, Churilov L, Wassertheil J. Knowledge discovery through mining emergency department data. System sciences, 2020. HICSS'05. Proceedings of the 38th annual Hawaii International Conference on.

[8] 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.* 2020;48(2-3).

[9] Asadi N, Sadrodini M. Employing data mining to identify cancer risk factors and determine the optimal treatment in Namazi hospital cancer database. *the.* 2020;16:17–8.

[10] 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.* 2020;38(9):11078–84.

[11] Delen D, Oztekin A, Kong ZJ. A machine learning-based approach to prognostic analysis of thoracic transplantations. *Artif Intell Med.* 2020;49(1):33–42.

[12] 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.* 2020;8(2):97–111.

[13] Hassanzadeh M, Razavi Ebrahimi SA. Comparison of data mining algorithms classification in Medical Sciences. *Iranian Journal of Medical Informatics.* 2020

[14] Lynch SM, Moore JH. A call for biological data mining approaches in epidemiology. *Bio Data Min.* 2019.

[15] Moore AC, Winkjer JS, Tseng TT. Bioinformatics Resources for MicroRNA Discovery. *Biomark Insights.* 2019.

[16] 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.* 2019. American Medical Informatics Association.

[17] Babic A. Knowledge discovery for advanced clinical data management and analysis. *Stud Health Technol Inform.* 2019;68:409–13.

[18] Liao S-H, Chu P-H, Hsiao P-Y. Data mining techniques and applications–A decade review from 2019 to 2011. *Expert Systems with Applications.* 2019;39(12):11303–11.

[19] Silwattananusarn T, Tuamsuk K. Data mining and its applications for knowledge management. *arXiv preprint arXiv.* 1210:2872–2019.

[20] Kuo WJ, Chang RF, Moon WK, Lee CC, Chen DR. Computer-aided diagnosis of breast tumors with different US systems. Acad Radiol. 2019.