International Journal of Computer Science and Software Engineering (IJCSSE), Volume 5, Issue 2, February 2016
ISSN (Online): 2409-4285 www.IJCSSE.org Page: 19-23

Clinical Decision Support Systems for Heart Disease Using Data Mining Approach

Harpreet Singh[1] and Kuldeep Singh Kaswan[2]

[1, 2] P.D.M College of Engineering, Bahadurgarh

ABSTRACT

Now a day's business is growing at a very rapid pace and a lot of information is generated. The more information we have, based on internal experiences or from external sources, the better our decisions would be. Business executives are faced with the same dilemmas when they make decisions. They need the best tools available to help them. Decision support system helps the managers to take better and quick decision by using historical and current data. By combining massive amounts of data with sophisticated analytical models and tools, and by making the system easy to use, they provide a much better source of information to use in the decision-making process. Health care is also one of the domains which get a lot of benefits and researches with the advent and progress in data mining. Data mining in medicine can resolve this problem and can provide promising results. It plays a vital role in extracting useful knowledge and making scientific decision for diagnosis and treatment of disease. Treatment records of millions of patients have been recorded and many tools and algorithms are applied to understand and analyze the data. Heart failure is a common disease which is difficult to diagnose. To aid physicians in diagnosing heart failure, a decision support system has been proposed. A classification based methods in health care is used to diagnose based on certain parameters to diagnosis if the patient have certain disease or not. The purpose is to explore the aspects of Clinical Decision Support Systems and to figure out the most optimal methodology that can be used in Clinical Decision Support Systems to provide the best solutions and diagnosis to medical problems.

Keywords: *Data Mining, Health Care, Heart Disease.*

1. INTRODUCTION

Data mining [1, 2] is concerned with finding patterns and models from the available data. Data mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Data mining or knowledge discovery in database, as it is also known, is the non-trivial extraction of implicit, previously unknown and potentially useful information from the data. This encompasses a number of technical approaches, such as clustering, data summarization, classification, finding dependency networks, analyzing changes, and detecting anomalies [3]. Data mining includes predictive data mining algorithms, which result in models that can be used for prediction and classification, and descriptive data mining algorithms for finding interesting patterns in the data, like associations, clusters and subgroups.

2. DATA MINING IN HEALTH CARE

Applications of data mining technology in medical domain include prediction of the effectiveness of surgical procedures and discovery of relationships among medicine and hospital data. In the last few years, the digital revolution has provided relatively inexpensive and available means to collect and store large amounts of patient data in databases containing rich medical information and made available through the Internet for Health services globally.

Data mining techniques applied on these databases discover relationships and patterns that are helpful in studying the progression of disease. Data mining plays an important role in medical diagnosis [1]. Neither medicine nor medical reasoning represents exact sciences, thus knowledge, which is hidden in patient records is valuable.

Information technologies in healthcare have enabled the creation of electronic patient records obtained from monitoring of the patient visits. This information includes patient demographics, records on the treatment progress, prescribed drugs, lab results, details of examination, , previous medical history, etc. Information system simplifies and automates the workflow of health care institution. Privacy of documentation and ethical use of information about patients is a major obstacle for data mining in medicine. In order for data mining to be more exact, it is necessary to make a considerable amount of documentation. Health records are private information, yet the use of these private documents may help in treating deadly diseases [4]. Before data mining process can begin, healthcare organizations must formulate a clear policy concerning privacy and security of patient records. The policy must be fully implemented in order to ensure patient privacy of data. Health institutions are able to use data mining applications for a variety of areas, such as doctors who use patterns by measuring clinical indicators, customer satisfaction and economic indicators, quality indicators, performance of physicians from multiple perspectives to optimize use of resources, cost efficiency and decision making based on evidence, optimize health care identifying high-risk patients and intervene proactively [5].

3. LITERATURE SURVEY

Although applying data mining is beneficial to healthcare, disease diagnosis, and treatment, few researches have investigated producing treatment plans for patients. Accurate diagnosis and treatment given to patients have been major issues highlighted in medical services.

Recently, researchers started investigating using data mining techniques to handle the error and complexity of treatment processes for healthcare providers. Razali et al. (2009) investigated generating treatment plans for acute upper respiratory infection disease patients using a decision tree. The model recommended treatment through giving drugs to patients showing accuracy of 94.73%. Applying association rules and decision tree to treatment plans are showing acceptable performance. However, the comparison with other data mining techniques such as naïve bayes, neural network, and genetic algorithms still needs investigation. Saad et al. (2010) presented the development of treatment plans to support treatment decision making for health care practitioners. The development of the treatment plan was generated on six common diseases using the decision trees technique showing an accuracy level that ranged from 77.97% to 91.67%. [8].

Although using data mining techniques in developing treatment plans for patients is showing acceptable performance, there has been little focus on treatment for heart disease patients. Kim et al. (2005) evaluated the current treatments for chronic heart failure using a decision tree and compared the results with those of large-scale clinical trials. They investigated which drugs can increase or decrease plasma level, fractional shortening, and left ventricular end-diastolic diameter in the cardiovascular disease. However, they did not investigate using data mining techniques to identify the suitable treatment for heart disease patients.

As hybrid data mining techniques showed promising results in the diagnosis of heart disease, so researchers investigating the use of hybridized data mining techniques in identifying the suitable treatment for heart disease patients are likely to be fruitful. The selection of a suitable treatment for heart disease patients is inherently complex, particularly as overall treatment frequently involves multiple, often concurrent, elements. This complexity makes treatment recommendation from data mining very difficult. Data mining is suited to assist decision making when many variables must be assessed, such as multiple concurrent treatments, but usually to make a single selection. To facilitate data mining being successful, our approach to analyzing treatment options and formulating recommendations using data mining will focus on what is the next best treatment step within the current treatment plan. Recognizing that this loses some richness from the overall treatment plan, it does make the problem more like one of classification, for which data mining is well-suited.

In light of the success that data mining techniques, and particularly hybridized techniques, have had in classifying heart disease sufferers, it seems important that similar approaches are considered in the selection of appropriate treatment for heart disease sufferers. From this, we propose a systematic approach to assessing the effectiveness of hybridized data mining techniques to identify the suitable treatment for heart disease patients.

4. METHODOLOGY

The experiment is based on Heart Disease data set. The data set having 14 attributes and 303 instances. The 14 attributes are age, sex, chest pain, blood pressure, cholesterol, fasting blood sugar<200, resting ecg, maximum heart rate, angina, peak, slope, colored vessels, thal, classs.

The objective is to find the promising classifier in order to diagnosis of heart disease among MLP, Random Forest, J48, ADTree.

Data Set Description

1. Age :29 to 77
2. sex: sex (1 = male; 0 = female)
3. cp: chest pain type
   -- Value 1: typical angina
   -- Value 2: atypical angina
   -- Value 3: non-anginal pain
   -- Value 4: asymptomatic
4. trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5. chol: serum cholestoral in mg/dl

6. fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. restecg: resting electrocardiographic results
   -- Value 0: normal
   -- Value 1: having ST-T wave abnormality (T wave inversions and/or T elevation or depression of > 0.05 mV)
8. thalach: maximum heart rate achieved
9. exang: exercise induced angina (1 = yes; 0 = no)
10. oldpeak = ST depression induced by exercise relative to rest
11. slope: the slope of the peak exercise ST segment
    -- Value 1: upsloping
    -- Value 2: flat
    -- Value 3: downsloping
12. ca: number of major vessels (0-3) colored by flourosopy
13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14 num: diagnosis of heart disease (angiographic disease status)
    -- Value 0: < 50% diameter narrowing
    -- Value 1: > 50% diameter narrowing

Table 1: Accuracy/Error Rate/Time

Classifier	Accuracy	Error Rate	Mean absolute error	Time Taken to build model
MLP	81.5	0.18	0.1972	1.75
Random Forest	72.93	0.270	0.27	0.03
J48	79.8	0.201	0.2323	0.09
ADTree	80.5	0.19	0.2467	0.09

Table 2: True Positive/False Positive/False Negative/True Negative

Model	TP	FP	FN	TN
MLP	195	24	32	52
Random Forest	179	40	42	42
J48	192	27	34	50
ADTree	191	28	31	53

Table 3: TPR/FPR/Precision/Recall/F-Measure/ROC Area of Healthy Class

Classifier	TPR	FPR	Precision	Recall	F-Measure	ROC Area
MLP	0.890	0.381	0.859	0.890	0.874	0.876
Random Forest	0.871	0.500	0.810	0.817	0.814	0.656
J48	0.877	0.405	0.850	0.877	0.863	0.717
ADTree	0.872	0.369	0.860	0.872	0.866	0.875

Table 4: TPR/FPR/Precision/Recall/F-Measure/ROC Area of Sick Class

Classifier	TPR	FPR	Precision	Recall	F-Measure	ROC Area
MLP	0.619	0.110	0.684	0.619	0.650	0.878
Random Forest	0.729	0.183	0.512	0.500	0.506	0.655
J48	0.595	0.123	0.649	0.595	0.621	0.717
ADTree	0.631	0.128	0.654	0.631	0.642	0.875

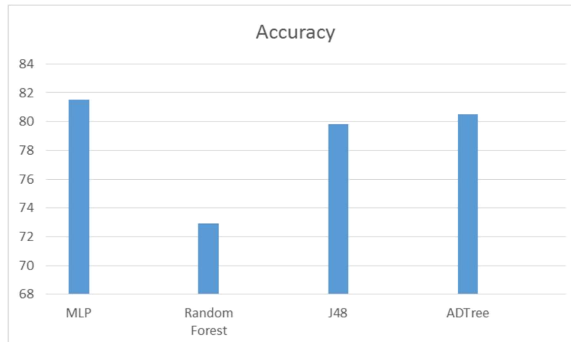

Fig. 1. Accuracy of Classifiers

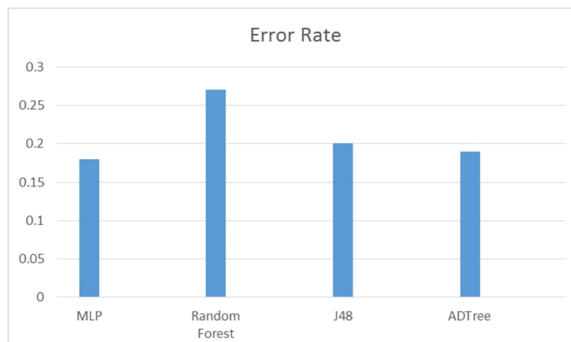

Fig. 2. Error Rate of the classifiers

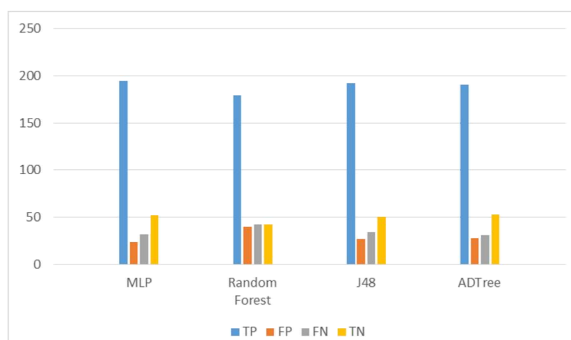

Fig. 3. True Positive/False Positive/False Negative/True Positive

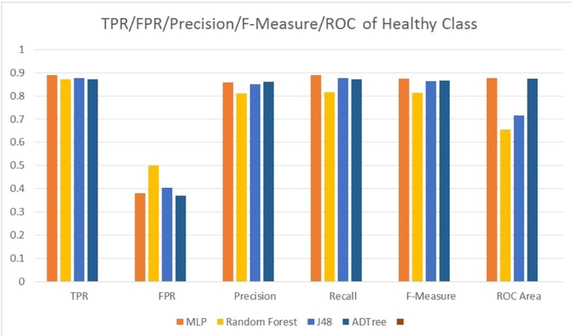

Fig. 4. TPR/FPR/Precision/Recall/F-Measure/ROC Area of Healthy Class

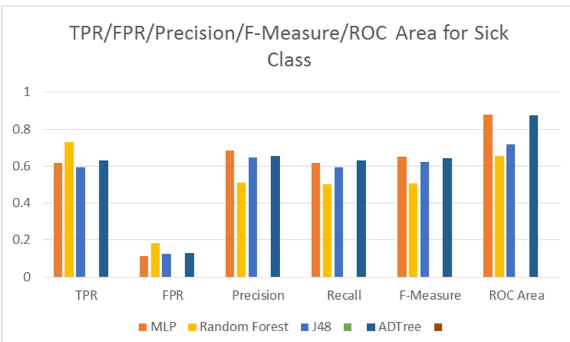

Fig. 5. TPR/FPR/Precision/Recall/F-Measure/ROC Area of Sick Class

5. OBSERVATIONS AND CONCLUSION

By analyzing the tables and figures result observations are as follows:

1. It is observed that error rate of Multilayer Perceptron (Table 1 is lowest i.e. –0.18 in comparison with classifiers, which is most desirable.
2. It is observed that Accuracy rate of Multilayer Perceptron (Table 1 is highest i.e.– 81.5 in comparison with other classifiers, which is most desirable
3. Value of True positive is high in case on Multiple Layer Perceptron.
4. It is observed that True positive Rate and True Negative rate of healthy class of Multiple Layer Perceptron is highest than other models .
5. It is also observed that MLP takes more time to build the model but other parameters to evaluate the model desirable for better classifier show that MLP gives accurate result as compared to other models
6. It is also observed that MLP takes more time to build the model but other parameters to evaluate the model desirable for better classifier show that MLP gives accurate result as compared to other models
7. It is observed that FPR of MLP .which is slightly lower than the FPR ADTree and as mentioned earlier that FPR is not main factor which determines the accuracy of binary classifier.

According to the experiments and result Multi Layer Perceptron gives a promising classification result for the tasks of classifying the cardiology data set with utmost accuracy rate and robustness.
In the present study, some of the works conducted on the field of Recommender Systems were studied. In addition, a number of clustering methods for clustering users in Recommender Systems were explained and separating grey-sheep users in the Recommender System was explicated. In addition, all methods tried to improve the quality of clustering users in Recommender Systems and also the reduction of mean absolute error.
Table 1 compares investigated methods and some of their features.
As suggestions for future research on clustering users in Recommender Systems, the following cases are illustrated:

1. The size of clusters has been considered fixed, in the future research, the size of clusters can be considered as one of the variables and can improve the precision of clustering.
2. Revolutionary algorithms such as genetic algorithm-based clustering and Imperialist Competitive Algorithm (ICA) were used for clustering users. These types of clustering causes that the possibility of the presence of normal users' errors in the cluster of grey-sheep users reduces.

REFERENCES

[1] Bushinak, H., AbdelGaber, S., & AlSharif, F. K. (2011). Recognizing The Electronic Medical Record Data From Unstructured Medical Data Using Visual Text Mining TechniquesProf. Hussain Bushinak. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 6 , 25-35
[2] Candelieri, A., Dolce, G., Riganello, F., & Sannita, W. G. (2011). Data Mining in Neurology. In Knowledge-Oriented Applications in Data Mining (pp. 261-276). InTech.
[3] Arun K Punjari, ―Data Mining Techniques‖, Universities (India) Press Private Limited,
[4] Canlas, R. D. (2009). Data Mining in Healthcare: Current Applications and Issues. Carnegie Mellon University, Australia.

[5] Koyuncugil, A. S., & Ozgulbas, N. (2010). Donor Research and Matching System Based on Data Mining in Organ Transplantation. J Med Syst (2010) 34 , 251–259

[6] Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines, international journal of medical informatics 77, 81–97.

[7] Houston, A. L., Chen, H., Hubbard, S. M., Schatz, B. R., Ng, T. D., Sewell, R. R., et al. (1999). Medical Data Mining on the Internet: Research on a Cancer Information System. Artificial Intelligence Review 13, 437–466

[8] Shouman.M et al, using data mining techniques in heart disease diagnosis and treatment, 2012 Japan-EgyptConference on Electronics, Communicationsand Computers

[9] Vrushali Y Kulkarni et al, “Random Forest Classifiers: A Survey and Future Research Directions”, International Journal of Advanced Computing, ISSN:2051-0845, Vol.36, Issue.1 April 2013.