International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 1 (2016) pp 139-142

Computer Aided Medical Decision Support: A Case Study on Severity Prediction of Toxic Epidermal Necrolysis

Manjusha. K K
Research Scholar, Department of Computer Science
Karpagam University, Coimbatore, Tamil Nadu, India.
E-mail: manjukrishna6002@gmail.com

K.Sankaranaryan
Professor, Department of Computer Science -SRIT
Coimbatore, Tamil Nadu , India.
E-mail: kkdcbesankar@gmail.com

Seena P
Asst. Professor, Department of Dermatology
Govt. Medical College, Kottayam, Kerala, India.
E-mail: drseenapalakkal@gmail.com

Abstract
A numbers of newer analytical tools have been introduced in the field of health care during the last few decades. Data mining is the one of the most intensively and extensively used technology to transform the complex and voluminous data gathered in health care organizations into useful information for decision making. Among the number of data mining algorithms and tools available our aims is to develop a decision support system with a graphical user interface for the prediction of severity of Toxic Epidermal Necrolysis (TEN) and also classifies the severity of disease into mild, moderate or severe by Naive Bayesian classification algorithm. The data was obtained from the tertiary health care centres of central Kerala, India. A study based on data collected during the period 2009 to 2014 was conducted. The study also compares the accuracy with J48 decision tree classification algorithm. The Naive Bayesian classification provides fast response and the accuracy being 97%. Accuracy with J48 for the prediction is 96%. The resultant confusion matrix indicates only few misclassifications in the data. The proposed decision support system shows accurate results in concurrence with the specialist's diagnosis. This study proves that naive bayes classification is very useful in the prediction of severity of TEN and helps the doctors to take assistance and to provide a better patient management. But computer aided diagnosis are not error free, so clinician's duty and dedications about patient care should never be misjudged, these systems helps to make their work easier.

Keywords: Decision support system, Naive Bayesian - classification, J48 decision tree, toxic epidermal necrolysis.

Introduction

Mathematical and statistical principles are widely used in the field of medicine as in prediction of cancer, heart diseases, diabetes etc. There are a lot of machine learning techniques used for it. Naive bayes classification is one of the multipurpose, robust statistical computational methodologies with firm theoretic back up and with strong potential to be effective in any discipline, especially in medicine. Naive bayes classification becomes significant in health care when getting data is very difficult especially in the case of rare diseases. Naive Bayes classifier require a small amount of training data to estimate the parameters essential for classification and it is suited when the dimensionality of the inputs is high [1].

Toxic epidermal necrolysis (TEN) is a rare, potentially deadly skin condition usually caused from drug reaction and can lead to discomfort and pain, and in some cases be fatal. The disease causes the top layer of skin to detach from the lower layer of the skin, all over the body. Any age group can find themselves affected by TEN, but normally seen in older patients. TEN diagnoses can be made clinically. The acute phase of TEN is characterized by persistent fever, severe mucous membrane involvement [2] Nikolsky's sign is almost present in toxic epidermal necrolysis. The main difficulties doctors have are the chances of misdiagnosis during the early stages of TEN. Skin diseases are mainly predicted by the doctors based on the appearance of the skin. But a number of parameters influence the prediction of such diseases. Prediction based only on appearance may lead to incorrect diagnosis. The proposed system predicts the severity of toxic epidermal necrolysis based on the parameters.

Data Mining Methods

Naive Bayesian Classification

Classification is an important research subject in machine learning approach. One of the important classification approaches is Bayesian classification. Naive Bayes classification is a simple, efficient and widely used Bayesian classifier. It is particularly appropriate when the number of parameters is high. A naive Bayes classifier is a simple probabilistic classifier based on Bayes's theorem with strong (naive) independence assumptions [3]. Bayes theorem calculates the probability factors over existing and proposed possible values over the samples. Bayes theorem can be represented by the simple notation

$$P(T|E) = \frac{P(E|T)XP(T)}{P(E|T)XP(T)+P(E|\sim T)XP(\sim T)}$$

T indicates the testing hypothesis and E represents the evidence. This is a conditional probability that is probability of one proposition is true provided that another proposition is true.

Decision tree

Decision trees are one of the most powerful tools in data mining and knowledge discovery. It has been used in analysis of large and complex bulk of data in order to discover useful patterns. The basic decision tree algorithm is called ID3 (Iterative Dichotomizer3). ID3 can handle only discrete values, but the successor C4.5 can handle numeric values. Classification and Regression Trees (CART) approach is suited for analysis of categorical and continuous datasets [4]. J48 is the implementation of ID3 algorithm developed by the WEKA [5]. It can handle different types of data like numeric, nominal, textual data and can also process incorrect or missing values. J48 can be implemented in data mining packages in different platforms and easy to understand because of its presentation. J48 show high performance with small effort.

Data Source

In this study, the data is collected from tertiary health care centres in the districts of Kottayam and Alappuzha, Kerala. For the selection parameters, we consulted dermatological practitioner from various medical colleges and were able to categorize the symptoms and ranges. With the help of these medical practitioners a questionnaire was prepared. These questionnaires were filled by doctors at the time of consultation. They also marked the improvement on the basis of mean- follow up of each patient. We used data collected during the period 2009-2014. A total of 110 records with 15 medical attributes were collected. The ethical committee approval was obtained. After the data cleaning process, (handling missing data, inconsistent data) we have 100 records with 15 factors for our proposed work. A data set of 100 patients was made, out of which 82 cases had varying severity of TEN and 18 cases reported dermatological presentations but had no evidence of TEN. On behalf of consistency, categorical and numerical attributes were used. The non categorical attributes included in the dataset were transformed to categorical data.

Table 1: Input attributes for the prediction of severity of toxic epidermal necrolysis

Input Attributes or Symptoms for Prediction		
1	Age	(value 0:<40; value 1:>40)
2	Anticonvulsant therapy	(value 0:Yes; value 1:No)
3	Eye symptoms	(value 0:redness of eye; value 1:Conjunctivitis; value 3:Corneal involvement)
4	Blood pressure	(value 0:>120; value1:120-100; value 2:<100)
5	Temperature	(value 0:≤100; value 1:100-102; value 2:>102)
6	Nikolsky sign	(value 0:positive; value 1:Negative)
7	Pulse Rate	(value 0:<90; value 1:90-120; value 2:>120)
8	Respiratory Rate	(value 0:<15; value 1:15-20; value 2:>20)
9	Urea	(value 0:20-40; value 1:40-50; value 2:>50)
10	Bicarbonate level	(value 0: upto 20; Value 1:20-30; value 2:>30)
11	Leukocytosis	(value 0:<10000; value 1:10000-14000; value 2:>14000
12	Area of involvement	(value 0:0-30; value 1:30-40; value 2:40-60; value 3:>60)
13	Skin Tenderness	(value 0:Mild; value 1:moderate; value 2:severe)
14	SGPT	(value 0:40-100; value 1:100-150; value 2:>150)
15	ALP	(value 0:140-150; value 1:150-200; value 3:>200)

Computer aided interpretation of medical data is widespread, now most of the physicians take the advice of computer applications to make their prediction accurate and reduce the time of diagnosis. So the result is, over the last few decades, a wide range of automatic systems has been developed in the area of medicine for decision support systems. The decision support system developed consists of three main parts: the input, the rule set and the output. A graphical user interface (GUI) was developed with Microsoft.Net 2010 with SQL Server as database and this GUI can be used by the physician. The data we collected and stored in the database can be used for future study and references. The doctor or user has to enter the patient parameter of TEN into the interface. The study predicted the severity of TEN, which is graded into mild, moderate or severe.

Implementation

The doctor or user can enter the symptoms of TEN into the graphical user interface for predicting the severity of patient. The input window is shown in figure 1. This GUI has been implemented in the C # .NET with Microsoft .Net 2010.

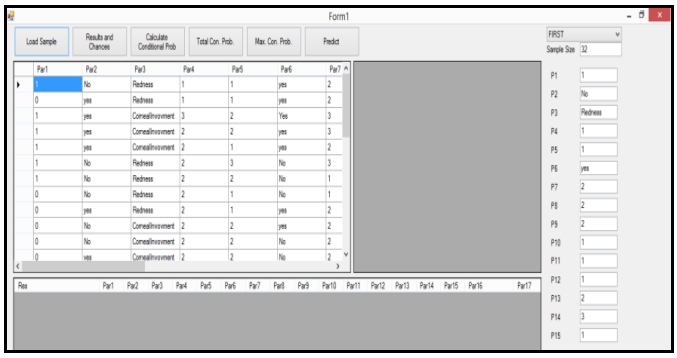

Figure 1: The graphical user interface showing the input window

A request from the user calculates the probabilities as per the below mentioned procedure.

P(CK)=P(A1|CK)*P(A2|CK)*P(A3|CK).....P(AN|CK).

Here P(CK) is the probability of a class CK and Ai is the set of all parameters (in our case, symptoms like age, fever, eye symptoms etc). N will be 15 parameters or symptoms. After finding the conditional probability we are finding the total conditional probability. Each step is included in the GUI. From the total probability, find the maximum conditional

probability. The result is inferred from the probabilities, which mean category with the highest probability is taken as the current status of the patient with the given symptoms. This formula will give the probability of the test symptoms to fall into one of the category, whether the severity is in mild, moderate or severe. Here we have

P(Mild) = .4375
P(Moderate) = .3125
P(Severe) = .03125
P(No) = .218758

From the above result it is clear that the patient with the symptoms falls into the Mild category as in figure 2.

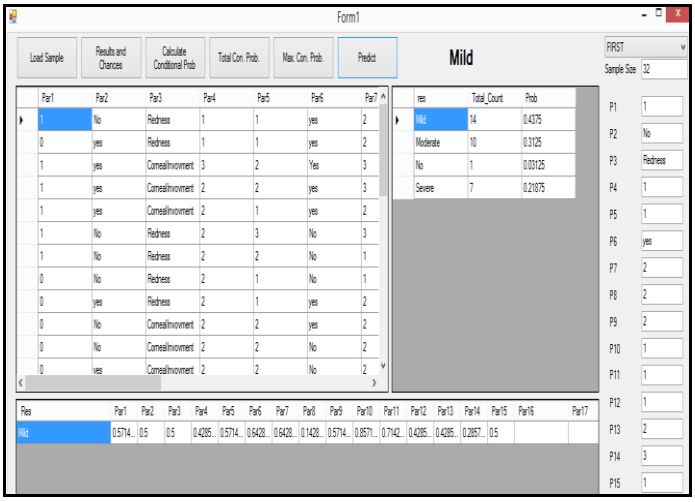

Figure 2: The graphical user interface of the Microsoft .Net program developed showing the results for a patient with milder form of toxic epidermal necrolysis.

Result

Performance of the expert system with graphical user interface for diagnosis of TEN patients is shown in Table 2. User can input the patient details into the interface and the result could be obtained. This software gave accurate results. The conditional probability of each parameter is calculated and stored in the database. A predicted percentage severity of less than 30% indicates cases of milder form. Similarly the values between 30 to 50 and greater than 50 indicate the moderate and severe forms respectively of the conditions. The results are in consistency with the type of treatment given to the patient and the days of hospitalization.

Naive Bayesian classify the disease with an accuracy of 97% and the decision tree was able to achieve an accuracy of 96%. The sensitivity, specificity and the precision were calculated. They are known in statistics as classification functions are statistical measures of the performance of a binary classification test. Sensitivity measures the proportion of actual positives which are correctly identified, i.e., it relates to the test's ability to identify positive results (e.g., the percentage of sick people who are correctly identified as having the condition). Specificity measures the proportion of negatives which are correctly identified, i.e., the ability of the test to identify negative results (e.g., the percentage of health people who are correctly identified as not having the disease [6]. The classification accuracies on the test are given in Table 2. The sensitivity, specificity and precision of the two data mining algorithms are compared.

Table 2: Accuracy comparison of the data mining models

Measure	Formula	Naive Bayesian %	Decision tree %
Recall (Sensitivity)	TP/(TP+FN)	96.8	96.2
Specificity	TN/(TN+FP)	98.8	98.8
Precision	TP/(TP+FP)	97.6	97.6
Accuracy	(TP+TN)/(TP +FN+FP+TN)	97	96

The proposed naive bayes model was able to classify 98.89% of the input instances correctly and decision tree 98.88%. Both algorithms exhibited good accuracy. Naive Bayesian algorithm showed more accuracy than decision tree algorithm.

Discussion

This is not the first automated system, many computer-assisted systems have been done earlier. MYCIN was one of the first rule-based expert systems in the clinical setting developed by Dr.Edward Shortliffe at Stanford University in 1970s [7]. MYCIN was used in the identification the type of bacteria causing infection. Later, ONCOCIN was developed in LISP in 1980's by Standford AI group was used in the treatment of cancer patients. The aim of the study was to evaluate the performance of naive bayes and J48 decision tree before selecting in disease diagnosis. From the above result naive bayes shows better results, so we came to the conclusion that naive bayes algorithm is best algorithm in the classification as compared with J48 for the prediction of severity of the disease. The program developed in our study with a graphical user interface built in c#.net was able to correctly diagnose the disease based on our input. This automated system enables the less experience junior doctors to arrive better diagnosis with the absence of skin specialist not to replace specialist. Specialist can also use this system for checking the progress of patients with their medicine.

Conclusion

Using computer aided techniques in medical field could reduce the medical error, time and cost. In this we can predict whether patient condition is severe, moderate or mild. So doctors can go through such treatment and also used to check the improvement of each patient condition. This model can further be improved by incorporating more parameters and the increasing the number of cases for training and testing. In this study a comparison using only two algorithms is done. A more number of algorithms can be tried like ANN or SMO for the comparison of data.

Reference

[1] Hardik Maniya, Mosin I. Hasan, Komal P. Patel. Comparative study of Naive Bayes Classifier and KNN

for Turberculosis. International Conference on Web Services Computing(ICWSC) 2011. Proceeding published by International Journal of Computing Applications.

[2] Bastuji-Garin S, Rzany B, Stern RS, Shear NH, Naldi L, Roujeau JC. Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. Archives of Dermatology. 1993 Jan; 129(1): 92-96.

[3] Bright TJ, Wong A, Dhurjathi R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems a systematic review. Ann Intern Med 2012; 157(1):29-43.

[4] Ho,T.J.: "Data Mining and Data Warehousing", Prentice Hall. 2005.

[5] www.data-mining.business-intelligence. uoc.edu/home/j48 - decision-tree

[6] Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ 1994; 308(6943); 1552.

[7] Shortliffe EH. Computer-based medical consultations: MYCIN. 388 ed. New York: Elsevier; 1976.