ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 09, pp.38328-38331, September, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

MULTIDISEASE PREDICTION AND TREATMENT ANALYSIS USING DATA MINING TECHNIQUES

***Anitha, S., Arulanandam, K. and Amudha Prabha, A.**

Department of Computer Science, Government Thirumagal Mills College, Gudiyattam-632 602, Vellore District, Tamilnadu, India

ARTICLE INFO

Article History:
Received 08th June, 2016
Received in revised form
29th July, 2016
Accepted 01st August, 2016
Published online 20th September, 2016

Key words:

Large Memory Storage and Retrieval (LAMSTAR),
Service oriented architecture (SOA),
WTA (Winner-Take-All).

ABSTRACT

Disease prediction and diagnosis is one of the complex applications where data mining tools and techniques are used to providing successful results because of significant improvements in technology. This research identifies gaps in the research on disease prediction, diagnosis and treatment and it also proposes a model to systematically close those gaps. Data mining have great potential for healthcare industry to enable health systems to systematically use data and identify the efficiency and improve care with reduce cost. The data mining techniques to Multi disease treatment it can provide reliable performance. So the system can be effective in reducing the death toll. The healthcare industry collects huge amounts of healthcare data which, unfortunately are not "mined" to discover hidden information for effective decision making. This proposed work has developed a prototype for the Multi Sickness Prediction System (MSPS) using data mining techniques by to compute the chance of prevalence of explicit unwellness from medical knowledge by using k-means, Large Memory Storage and Retrieval (LAMSTAR) and Medical diagnosis methodology. The system uses service oriented architecture (SOA) whereby the system elements of diagnosis, data portal and alternative miscellaneous services are provided. This reduces the multiple diseases showing the similar symptoms problem and it will increase the accuracy of such diagnosis. This proposed system will provide some reliable decision finding the disease for healthcare support.

Citation: Anitha, S., Arulanandam, K. and Amudha Prabha, A. 2016. "Multidisease prediction and treatment analysis using data mining techniques", *International Journal of Current Research*, 8, (09), 38328-38331.

INTRODUCTION

Patients can receive better more affordable healthcare services with the best curable treatment for the particular disorder. Sickness diagnosis and prediction involves multiple physicians from different specializations in case of cancer, liver disorders and heart disease. This requires multiple biomedical markers and multiple clinical factors like the age, general health of the patient, its location, and type of disease, the grade and size of the disorder. For reasonable prediction information like cell based, patient based and population based all must be carefully considered by the attending medical practitioner. It is challenging even for the most skilled technician to do. Both physicians and patients need to face same challenges when it comes to the matter of disease prevention and disease prediction. Sometimes these conventional clinical, behavioural parameters and environment may not be sufficient to make better predictions.

****Corresponding author: Anitha, S.***
Department of Computer Science, Government Thirumagal Mills College, Gudiyattam-632 602, Vellore District, Tamilnadu, India

In most of the critical cases to predict the disease need some specific molecular details about either the infected part or the patient's genetic status. With the speedy development of the proteomic, genomic and imaging technologies, this molecular scale information about patients is now can be readily acquired.

Architectural Model

In architectural model it contains two databases: Patient Records database and Disease/Symptoms database. Four web services are used to implement the SOA. They are Pattern matching, recent trends, differential diagnosis and recent differential diagnosis. The patient Record database contains all the patient information from all the hospitals in the network. Diseases/Symptoms database is a centralized database. First the doctor retrieves the symptoms from the patient record database. After retrieving the symptoms, the doctor identify whether any symptom related diseases contains in the Diseases/Symptoms database. Here the pattern matching service is activated. If any diseases match with Symptoms means list out all the possible matched symptoms and presents the result to the doctors. If the doctors not satisfied with

results, compare to recent history and recent trend service must be activated. This service makes use of the Diseases/symptoms database and Patient Record database and the result obtained from pattern matching service to get results.

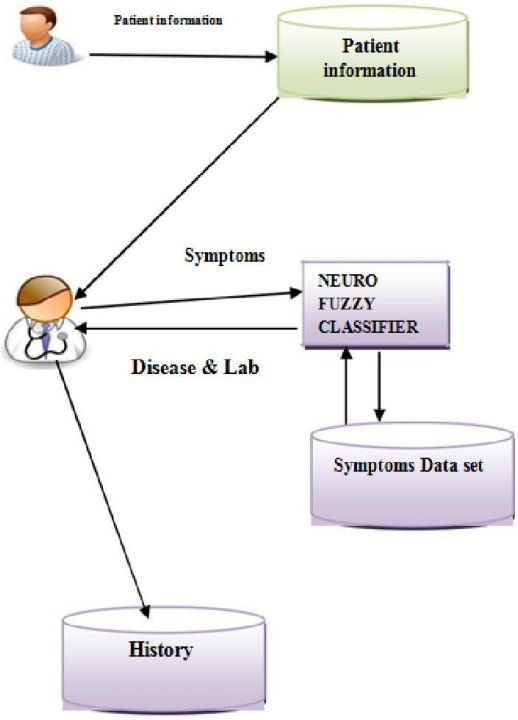

Fig.1. Architectural Model

Proposed Approach

The main motivation in this research is based on the assumption that the instance with similar attribute values is more likely to have similar class label. Similarity is measured based on Euclidean distance The reason in choosing k-means is that Lange *et al*. (2004) proved that the validation result obtained by k-means clustering is better the produce the effective aforesaid services and would increase the response time of the system. This approach tried clustering by k-mediod algorithm but the misclassification rate was 50%.

Step 1: Configure the dataset 1-m-n
Where 1=no of input, m – no of hidden inputs, n - no of output values.
Step 2: The no of weights are calculated based on storing Table W.
Step3: Assume no of digits in weight from Dataset D.
Step4: Choose symptoms Si form population Dataset pi,
Step 5: for each weighted symptoms
{
Extract weight W-I;
Keep the weight for each input for train Dataset.
Calculate the fitness Value Fi for each of symptoms from population dataset;
}
Step 6: apply LAMSTAR
Step 7: output Fi for each Si;

Fig. 2. Prediction Algorithm

The main advantage of this algorithm is simplicity and its speed which allows running large datasets. K-Means may be faster than hierarchical clustering (if K is small). K-Means may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

The lamstar neural network

The LAMSTAR (Large Memory Storage and Retrieval) is a Neural Network. The network employs standard perceptron-like neurons that are arranged in many SOM (Self-Organizing Maps) Kohonen modules (layers). Their SOM structure implies that their neurons are WTA (Winner-Take-All) neurons whose memory is stored in BAM-fashion (Bidirectional Associative Memory). However, the LAMSTAR network differs from most neural networks in that a key feature of the LAMSTAR is its employment of link-weights between the neurons of the various SOM modules and between these and neurons of SOM-type output modules. Hence, learning takes place both in the setting of memory-storage weights and in the setting of link-weight matrices (inter-relation or correlation-weights or Verbindungen, in Kantian terms). Decisions are therefore based, not on the memory values, but both on the memory elements (stored values) and on the connections (relations) between memory elements. Linkweights are learnt by reinforcement, in a Hebbian manner.

Retrieval of Information in the LAMSTAR Network

In applications such as medical diagnosis, the LAMSTAR system is trained by entering the symptoms/diagnosis pairs (or diagnosis/medication pairs). The training input vectors **X** are of the following form:

$$\underline{X} = \left[\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_N^T\right]^T$$

The LAMSTAR NN was shown to be very fast in its computation due to its employment of link-weights when combined with winner-take-all associative memory, in a many-layer structure. The prime objective of the proposed algorithm is to defeat the difficulty in FNN algorithm that is sensitive to initial condition. Selecting the different initial condition may attain different cluster results. The algorithm may be caught in the local optimum. For the same medical dataset, different cluster may lead to different sets of rule for building a classifier model. The accuracy of the system again depends on the local optimum.

Neuro Fuzzy classifier

The training was carried out with a random set of initial values for the premise and consequent parameters and then manually tuning them. The training process was carried out with fifty different set of initial values. The process of training the network was stopped, for every initial set of values, when the improvement in the classification accuracy was not varying for at least fifteen consecutive runs. The time taken to perform the test runs was one man month. The knowledge (ai, bi, pi, qi and ri) gained by the trained network is stored in the knowledge base. The Neuro Fuzzy Classifier takes two features x and y as

input. The features of hepatitis data were reduced to x and y by applying PCA and FCM. The Boolean output z, of this classifier indicates whether the patient will the disease (TRUE) or not (FALSE). Model tailored using neuro-fuzzy inferencing technique for predicting diseases shown in Fig. 3.

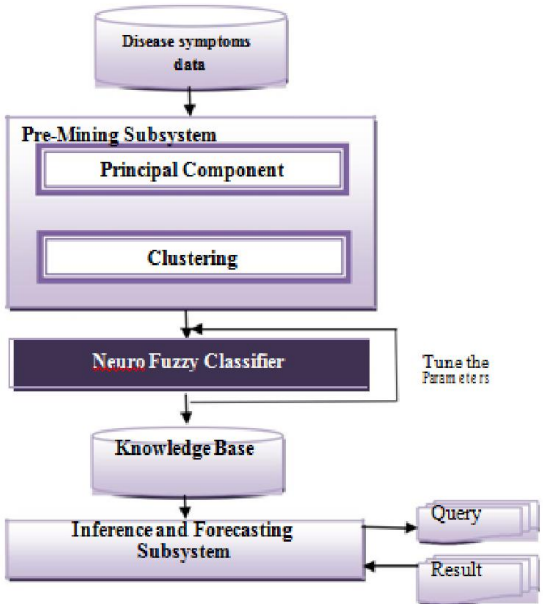

Fig.3. Model Tailored Using Neuro-Fuzzy Inferencing Technique

Symptoms comparing using iterative serach

In this phase symptom matching using iterative search utilize data that is stored. The first step of the algorithm involves selecting the symptoms shown by the patient. The algorithm gives the list of all possible diseases ranked according to the number of symptoms matched in the database. The list is generated after input of every symptom. After the first iteration for the second iteration the next list of symptoms will be shortlisted according to the disease list that was obtained in the previous iteration. The new symptom list will contain symptoms of only those diseases that were obtained in the previous list, if *headache*, *fever* and *pain* in the *sinuses* are entered, then the weights W15, W16 and W19 will be considered. Next all the weights will be added and compared to all subclasses C1, C2, C3 and C4 is most likely the answer depending on its weight. Finally all the diseases in class C4 are considered and if sinusitis (D4) weight is closer to the sum of all the input symptoms weights, then it is possible diagnosis.

RESULTS

In the current research work, identify the disease based on the patient symptoms are estimated using data mining classification techniques. The proposed According to Neuro Fuzzy Classifier (NCN) and k – means algorithm was able to classify 94% of the input instances correctly. It exhibited a precision of 91% on an average, recall of 86% on an average, and F-measure of 91.2% on an average. This inference system prepared a confidence measure which gave the probability of correct suggestions by examining the values with the inference calculation. This was done by calculating the residual value. Neuro Fuzzy Classifier was used to calculate the membership probability of a given patient record. The graph showing the comparison of execution time with the number of disease prediction is given in Figure 4.

Table 1. Sample Data Set: Iterative Pattern Search

Disease	Symptoms and weight	Class weight
AIDS	Fatigue, swollen lymph nodes, ulcers in the mouth or on the genitals, muscle aches and joint pain, nausea and vomiting, night sweats, body rash, fever	A1 Cd4 Cells
Dengue Fever	high fever, Severe headaches Pain behind the eyes, Severe joint and muscle pain, Fatigue, Nausea, Vomiting, Skin rash	D1
H1N1 flu	Fever (but not always), Cough, Sore throat, Runny or stuffy nose, Watery, red eyes, Body aches, Headache, Fatigue, Diarrhea, Nausea and vomiting	W1

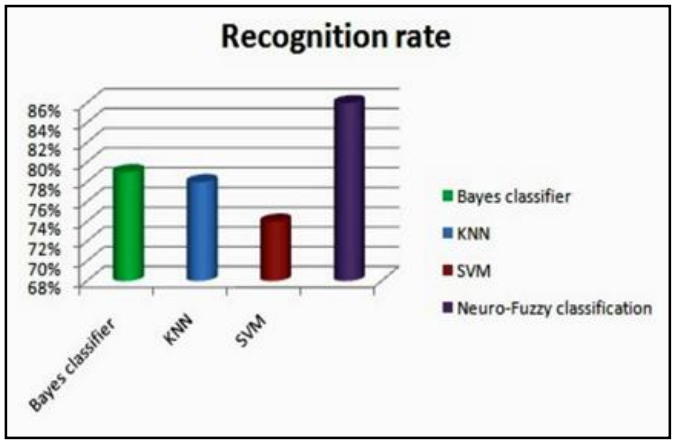

Fig. 4. Recognition Rates of Classification Algorithms

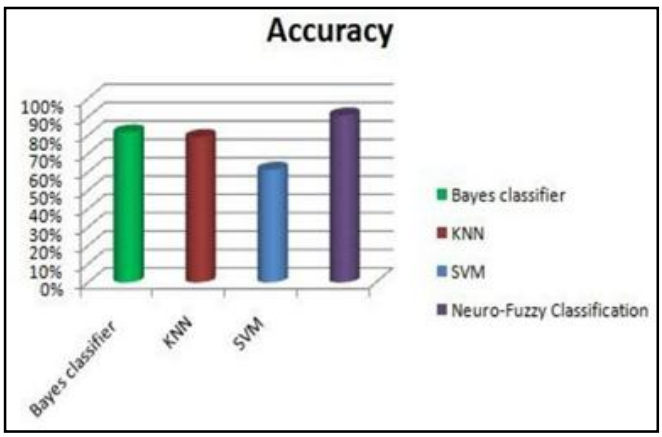

Fig. 5. Accuracy comparison between the classification algorithms

The best classifier based on a k-means clustering algorithm reached recognition rates above 86 % in comparison to the Bayes classifier (79 %) and the KNN classifier (78 %). These results suggest that Neuro-Fuzzy algorithms have the potential to significantly improve common classification methods for the use in disease prediction. The results of all Four models, Neuro Fuzzy Classifier appears to be most effective as it has

the highest percentage of correct predictions (90.79%) for patients symptoms, followed by followed by naïve Bayes, support vector machine and KNN. The performance in terms of graphs for accuracy, precision, sensitivity, specificity. Figure 5 shows the accuracy percentage obtained by high risk and low risk. With the use of the information gain as split parameter in Neuro Fuzzy Classifier, the results are exhibited by average precision, recall and accuracy of this technique was found to be 90.79 %. Niyati Gupta *et al.* (2011) have defined the accuracy as the proportion of instances that are correctly classified. It is calculated by the total number of correctly predicted "high risk" (true positive) and correctly predicted "low risk" (true negative*)* over the total number of classifications. It can be calculated as

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Conclusion and future work

Health care relevant data are enormous in nature and they arrive from various birthplaces all of them not wholly relevant in structure or quality. These days, the performance of knowledge, observation of various specialists and medicinal screening data of patients grouped in a database during the analysis process, has been widely accepted. In this Research thesis is presented an efficient approach for Multi disease prediction based on the patient symptoms warehouses for the efficient prediction of diseases. This research uses, LAMSTAR Network, K-Means algorithm and Neuro Fuzzy Classifier to assist the doctors to perform differential diagnosis along with the possible implementation using SOA technique. By using these techniques, it improves the overall speed and increase the accuracy of algorithm. Especially in large datasets, LAMSTAR network gave faster and better result. It reduces the effects of misdiagnosis, especially practioners and students can also easily identify the diseases. The results obtained for the prediction of the multiple disease show that the system can classify the positive samples with better accuracy as compared to classification of negative samples classification. It can be observed that the classification accuracy of the neuro-fuzzy approach for hepatitis data is relatively better when compared to the other approaches that use neural networks with back propagation training.

It will also help the medical fraternity in the long run by helping them in getting accurate diagnosis and sharing of medical practices which will facilitate faster research and save many lives. In future work, this research will extend to conduct experiments on large real time health datasets to predict the diseases and compare the performance of this algorithm with other related data mining algorithms. In future this work is to be extended by using other data mining algorithms and optimization algorithms for predicting other diseases from the hemogram blood test data set by using reduced attributes.

REFERENCES

Akhil Jabbar, M., Deekshatulu, B.L. and Priti Chandra, 2013. "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm*"*, International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013.

Nadali, A., Kakhky, E.N., Nosratabadi, H.E. 2011. "Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system," Electronics Computer Technology (ICECT), 2011 3rd International Conference on, vol.6, no., pp.161, 165, 8-10 April 2011.

Patil, B.M., Ramesh C. Joshi and Durga Toshniwal, 2010. *"*Effective framework for prediction of disease outcome using medical datasets: clustering and classification*" Int. J. Computational Intelligence Studie.* 2010.

Rahul Isola, Rebeck Carvalho and Amiya Kumar Tripathy. 2012. "Knowledge discovery in Medical system by using Differential Diagnosis, AMSTAR and K-NN, IEEE Transaction on Information Technology in Biomedicine," Vol.16, No.6, November.

Shweta Kharya, 2012. *"*Using Data Mining Techniques For Diagnosis and Prognosis of Cancer Disease*", International Journal of Computer Science, Engineering and Information Technology* (IJCSEIT), Vol.2, No.2, April 2012.

****