AN EXPERT SYSTEM TO DETECT SYSTEMIC LUPUS ERYTHEMATOSUS DISEASE USING APRIORI ALGORITHM AND NEURAL NETWORK ALGORITHM – A DATA MINING APPROACH

Gomathi. S[1], Dr. V. Narayani[2]

[1]Research Scholar, Bharathiar University, IT Department, SKASC, Coimbatore, (India)

[2]Director I/C, Department of MCA, Karpagam College of Engineering, Coimbatore, (India)

ABSTRACT

The research work focuses on developing an expert diagnosis system to detect Systemic Lupus Erythematosus (SLE) disease using association rule and neural network. Neural network is a classification algorithm and an efficient way to classify the lupus data set and association rule is used to reduce the complexity and dimensions of disease repository. The classification algorithm is compared with two algorithms (J48 and Support Vector Machine) to measure the performance of neural network. Apriori algorithm (Association rule algorithm) is used with neural network to obtain the fastest, efficient and complex free diagnostic system to detect SLE disease. A new framework is proposed to detect the disease in earlier stage.

Keywords: Apriori; Neural Network; Antibodies; Arthritis; Serositis; Vasculitis

I. INTRODUCTION

SLE is an auto immune inflammatory and chronic disease which produces auto antibodies on white blood cells (WBC). It affects females than male in the ratio of 10:1 and between 20-40 age limit [4]. It affects many organs so called as multi system disorder. Many organs include cardiovascular, pericarditis, coronary artery, myocarditis, skin, lungs, kidney nervous system etc [1]. There are no curing methods or surgeries for Lupus but it can be predicted in advance if advanced technique has been implemented. Since the disease affects all parts of the body, it predicts like other disease. Thus it is difficult to predict in advance with the normal prediction methodology. Recent study shows that 10% of patient approaches the doctors after 5 years of lupus without knowing that they have lupus disease. The pathogenesis of lupus are environmental, family history, stress etc [3]. To predict lupus, huge data about patient is necessary which involves family history, age, gender, ethnicity, lab results etc. The data may be misinterpreted somewhere or mining the hidden data and pattern may be difficult. Thus data mining technique is used to mine the hidden pattern and utilization of knowledge can be done efficiently using data mining methodologies.

An expert system is developed using data mining classification and association rule technique. Association rule is used to find the relationship among voluminous datasets and Neural network classification algorithm is used to classify and predict the disease easily and efficiently.
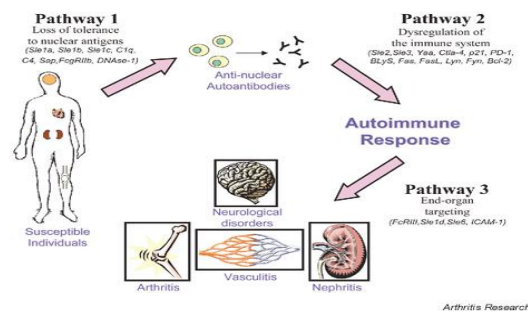
Fig. 1. Occurrence of Lupus

II. LITERATURE REVIEW

A deep survey about data mining techniques like association rules, clustering, classification, sequential patterns and prediction was done by Vijayarani and Sudha [2]. The techniques are applied for all the fields like health care, banking, agriculture, fraud detection etc. The main area concentrated using data mining is heart disease, diabetes, breast cancer etc. The authors defined an association rule algorithm for medical data set. The algorithm is based on three tasks (i) categorical and numeric attributes are initially transformed to transaction dataset, (ii) finding the predictive association rules with related attributes and (iii) validation of training and testing approaches. Various algorithms like Genetic algorithm, fuzzy learning, decision tree, C4.5, back propogation, Artificial Neural network, KNN, Naive Bayes theorem are depicted along with their applications. Naive Bayes shows the best accuracy level.

Usharani [5] did a parallel approach using neural network technique to diagnose breast cancer. Feed forward neural network and back propagation algorithm was used for training set. The performance was evaluated. Neural network can be efficiently implemented for biological models and considered as a powerful technique for data analysis. Neural network parallelism can be implemented in three ways exemplar parallelism, block parallelism and neuron parallelism. The three main aspects used by the author in constructing neural networks are structure, encoding and recall. The two passes in feed forward network in back propagation are forward pass and backward pass. Author analyzed 699 instances of breast cancer data set and used 10 attributes. The performance of single layer and multilayer neural network has been tabulated.

Usharani [6] classified the heart disease data set using neural network approach. Processing elements, neurons and weighted interconnection among points are important to form neural network. Neural network was used for fault tolerance and can be divided based on supervised training and unsupervised training. Artificial Neural Network is an iterative process. Author analysed 414 instances with 13 main attributes. Back propagation with momentum and variable learning rate was used for network training.

Authors [7] used cart algorithm to monitor diabetes. A new platform was designed for enhanced health monitoring for early detection of patient. Diabetic diagnostic and monitoring system used as a layered approach which provide decision supporting functionality for doctors and also provide easy access. CART shows 96.1% accuracy. The key element of CART is splitting each node in a tree, deciding when a tree is completed and assigned each terminal node to a class. The process of CART is pruning, testing and cross validation. CART will automatically search for important patterns and relationships and then quickly uncovers hidden structure. CART will split the data set according to a Cartesian that maximizes the separation of class with tree like structure.

Gal Yaniv et.al., [9] publicized the effect of Systemic Lupus Erythematosus in central nervous system (CNS). Imaging modalities of brain was considered in this paper. The modalities are categorized as morphological and functional. Magnetic Resonance Imaging (MRI), Diffusion-Weighted Imaging (DWI), Diffusion Tensor Imaging (DTI) and Magnetic Resonance Spectroscopy (MRS) are the important imaging results to assess lupus in CNS, its patho physiology and consequences. The imaging reports will show the functional view as well. SPECT scan also used to predict and assess the lupus disease.

III. PROBLEM SPECIFICATION

Hospital Information system are not centralized in India. Some hospitals use decision support system with limited facilities. If the lupus patient initially suffered with skin disease, the patient may approach dermatologist and his details will be recorded there. After few months the same patient may be affected with heart problem and the patients will consult with the cardiologist. The cardiologist don't know about the history of the patient. Like this the disease will occur one after another and the patient will meet the specialised doctor. After 4 years the patient will approach a high multispecialty hospital where he will have all kind of tests and complete treatment. Sometimes the data may be mis-interpreted. Finally the patient will be diagnosed with the disease later and he/she dies earlier. This needs a decision support system to predict the disease earlier and to extend the life time of the patient.

IV. PROPOSED WORK

4.1 Apriori Algorithm

Apriori algorithm is an iterative algorithm which finds the set of large 1 item set and set of large 2 itemset and so on. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time known as candidate generation, and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found [8].

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon)\\
&\quad L_1 \leftarrow \{\text{large } 1 - \text{itemsets}\}\\
&\quad k \leftarrow 2\\
&\quad \textbf{while } L_{k-1} \neq \emptyset\\
&\qquad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}\\
&\qquad \textbf{for } \text{transactions } t \in T\\
&\qquad\quad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}\\
&\qquad\quad \textbf{for } \text{candidates } c \in C_t\\
&\qquad\qquad count[c] \leftarrow count[c] + 1\\
&\qquad L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}\\
&\qquad k \leftarrow k + 1\\
&\quad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k-1. Then it prunes the candidates which have an

infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

The pseudo code for the algorithm is given for a transaction database T, and a support threshold of ϵ. Usual set theoretic notation is employed; though note that T is a multi set. C_k is the candidate set for level k. At each step, the algorithm is assumed to generate the candidate sets from the large item sets of the preceding level, hiding the downward closure lemma. Count [c] accesses a field of the data structure that represents candidate set c, which is initially assumed to be zero.

4.2 Neural Network

Neural Network is distributed information processing structure which has multiple number of processing elements denoted as node, which in turn interconnected via unidirectional signal channels called connections. There are two main classification of Neural Network called Supervised and unsupervised learning. Multi Layer Neural Network is used with back propagation network. Back propagation elucidates the difference between actual and predicted values which is propagated from output nodes backwards to nodes in previous layer.

Purpose of Choosing NN: The main purpose of choosing Neural network is fault tolerance, evidential response, adaptivity, learning ability, input-output mapping, non linearity and neurological analogy.
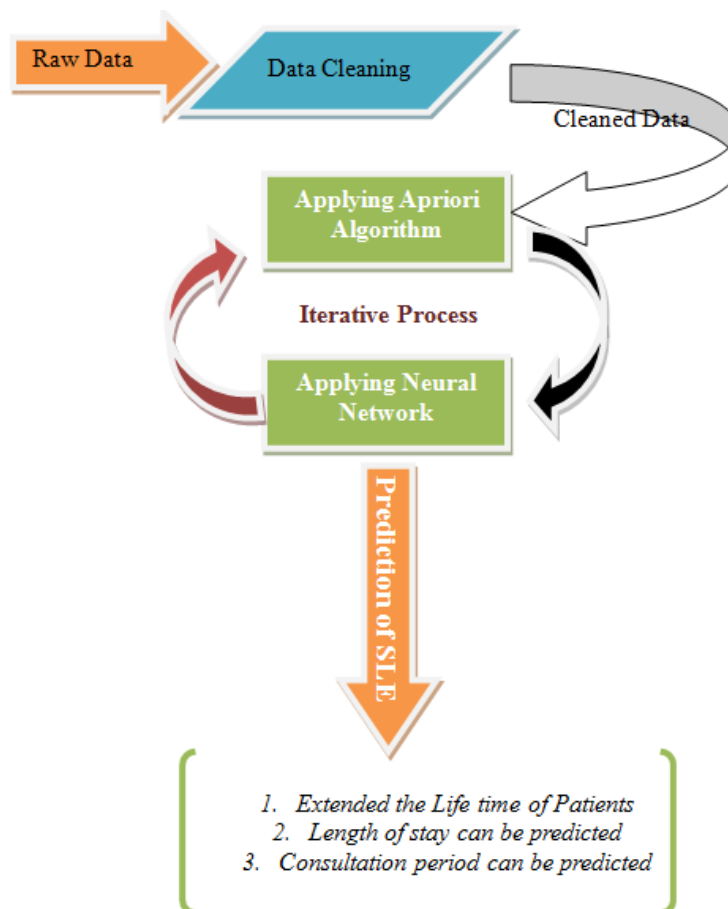
4.3 Expert System

Fig. 2. Expert System to Predict Lupus

The patient's raw data is gathered and the data is cleaned in the proper format. Once the data is cleaned, it is grouped and related using Apriori algorithm. The output of Apriori algorithm is given as the input of Neural Network. This is the iterative process till the proper prediction has been done.

4.4 Comparision of Other Algorithms

Table. 1. Comparison of Accuracy and Sensitivity

Other Algorithms Applied wih Apriori algorithm	Sensitivity	Accuracy
Neural Network	98.5	99.25
J4.8	98	98.5
ID3	97.5	98.5

The Sensitivity and accuracy of J4.8, ID3 and Neural Network when applied with Apriori is shown in Table 1. It is clear that Neural Network shows the highest and best sensitivity and accuracy when applied with Apiori algorithm.

V. CONCLUSION

An expert system to predict lupus is presented in this paper using Apriori and Neural Network Algorithm. Neural Network proves the better output which helps the domain experts for better diagnosis and results Comparison with other classification technique is done. The future work will be to apply this expert system to 100 patients and check whether the system is working effectively or not.

REFERENCES

[1]. A.T. Sayad, P.P. Halkarnikar, Risk level prediction system in ischaemic heart disease", Proceedings of IRF International conference, 30th march 2014. ISBN: 978-93-82702-69-6.

[2]. Vijiyarani, S, and S Sudha. "Disease prediction in data mining technique–a survey", International Journal of Computer Applications & Information Technology, Vol II, Issue I Jan p 17-21 ISSN 2278-7720.

[3]. Chauraisa, V, and Saurabh Pal. "Early Prediction of Heart Diseases Using Data Mining Techniques." Carib. j. SciTech Vol 1 2013 p 208-217.

[4]. V. Manikantan and S. Latha, "Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods", International Journal on Advanced Computer Theory and Engineering (IJACTE), Vol 2, Issue 2, 2013, ISSN (Print) : 2319 – 2526.

[5]. Rani, K Usha. "Parallel approach for diagnosis of breast cancer using neural network technique." International Journal of Computer Applications Vol 10 issue 3 (2010): p 1-5.

[6]. Rani, K Usha. "Analysis of heart diseases dataset using neural network approach." International Journal of Data Mining and knowledge management process, Vol 1, No 5, Sep (2011).

[7]. Kavitha, K, and RM Sarojamma. "Monitoring of Diabetes with Data Mining via CART Method." International Journal of Emerging Technology and Advanced Engineering Vol 2 Issue 11 (2012): p 157-162.

[8]. Chanchal Yadav, Shuliang Wang, Manoj Jumar, "Algorithm and approaches to handle large Data- A Survey.", International Journal of Computer Science and Network, Vol2, Issue 3, March 2013 http://arxiv.org/pdf/1307.543.

[9]. Gal, Y, Gilad Twig, Oshry Mozes, Gahl Greenberg, Chen Hoffman, Yehuda Shoenfeld, "Central nervous system involvement in systemic lupus erythematosus: an imaging challenge." The Israel Medical Association journal: IMAJ 15.7 (2013): 382-386.