DOI: 10.21276/sjet

Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2017; 5(8):427-439 **ISSN 2347-9523 (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Research Article

An Effective Diagnosis of Pulmonary Tuberculosis using K-Means Clustering and ANFIS

Dr. B. Ashadevi[1], Prof. P. Muthamil Selvi [2], B. Sasi Revathi[3]
[1,2]Assistant Professor / Computer Science, M.V. Muthaiah Govt. Arts College for Women Dindigul, India
[3]Sree Hayagreeva Arts and Science College, Dindigul, Tamil Nadu, India

***Corresponding author**
Dr. B. Ashadevi
Email: asharajish2005@gmail.com

Abstract: Data mining is the process of automatically extracting knowledgeable information from huge amounts of data. It has become increasingly important as real life data enormously increasing. Data mining is an integral part of KDD, which consists of series of transformation steps from preprocessing of data to post processing of data mining results. The basic functionality of data mining involves classification, association and clustering. Classification is a pervasive problem that encompasses many diverse applications. To improve medical decision data mining techniques have been applied to variety of medical domains. A major challenge that many of the health care organizations are facing is the provision for lack of quality services like diagnosing patients correctly and administering treatment at reasonable costs. Data mining techniques answer several important and critical questions related to health care. We propose an approach to predict the heart diseases using data mining techniques. In this paper, we investigate on K-Means Clustering and illustrate a reliable prediction methodology to diagnose tuberculosis disease and classify between different stages of tuberculosis using Adaptive Neuro Fuzzy Inference System (ANFIS). This prediction model helps the doctors in efficient heart disease diagnosis process with fewer attributes. Heart disease is the most common contributor of mortality in India.
Keywords: Data mining, KDD, k-means clustering, fuzzy sets, Neuro fuzzy logic

INTRODUCTION

While large scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open ended user queries. Several types of analytical software are available: Statistical, machine learning, and neural networks.

Pulmonary Tuberculosis

Tuberculosis (TB) is caused by infection with Mycobacterium tuberculosis, which is transmitted through inhalation of aerosolized droplets. TB mainly attacks the lungs, but can also affect other parts of the body (extra pulmonary tuberculosis).The disease is among the leading causes of mortality in India. India accounts for 1/5 of the global TB burden. Practitioners should identify all pulmonary tuberculosis suspects and get their sputum tested from a quality assured microscopy center. Under the Revised National Tuberculosis Control Program (RNTCP) more than 13,000 such quality-assured microscopy centers are available across the country wherein sputum sample may be sent for examination. Two sputum samples (one sample preferably early morning sample) need to be sent to quality-assured microscopy center. A patient with one or two sputum samples being positive for acid fast bacilli (AFB) by direct microscopy is diagnosed as having smear positive.

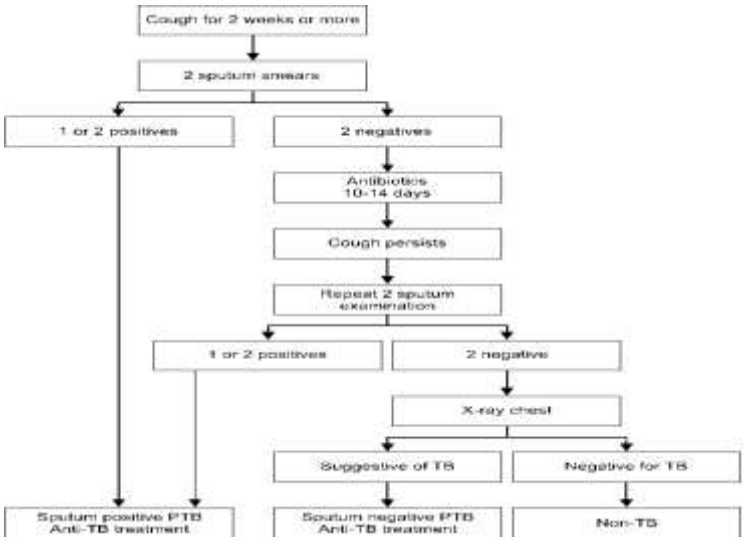

Fig-1: Flowchart for diagnosis of Pulmonary TB

Drug-resistant Tuberculosis

The diagnosis of drug resistant tuberculosis is laboratory based from quality assured, culture and drug susceptibility testing (C & DST) laboratory. Under RNTCP, 43 quality-assured C & DST laboratories are available across the country for diagnosis. Following categories of patients are considered as multidrug resistant tuberculosis (MDRTB) suspects: all patients who have failed first line treatment, all previously treated patients; all HIV-TB co-infected patients, any smear positive follow-up new or previously treated patients and all pulmonary tuberculosis cases who are contacts of MDR-TB.

With the worldwide re-emergence of TB, multi-drug resistant (MDR-TB), extensively drug resistant (XDR-TB) and extremely drug resistant strains have become an even greater threat. According to the WHO Global Tuberculosis Control Report 2009, there may be more than 500000 cases of MDR-TB worldwide. Current testing for drug resistance can take more than 4 weeks, leading to higher mortality and the further spread of MDR strains. 1.3. Treatment

The goal of treatment of tuberculosis is to ensure high cure rates, prevent emergence of drug resistance, minimize relapses and cut the chain of transmission through early diagnosis and treatment. TB can be treated effectively by using first line drugs (FLD) isoniazid (INH), rifampin(RIF),pyrazinamide (PZA), ethambutol (EMB) and streptomycin (SM). However, this first line therapy often fails to cure TB for several reasons. Relapse and the spread of the disease contribute to the emergence of drug resistant bacteria. The emergence of multidrug resistant TB (MDR-TB), i.e. which is resistant to at least isoniazid (INH) and rifampicin (RIF), is of great concern, because it requires the use of second-line drugs that are difficult to procure and are much more toxic and expensive than FLDs [3].Therefore, the detection and treatment of drug susceptible or single drug resistant TB is an important strategy for preventing the emergence of MDR-TB [6].

M. tuberculosis strains with extensively drug resistant-TB (XDR-TB), that is resistant to either isoniazid or rifampicin (like MDR tuberculosis), any fluoroquinolone, and at least one of three second-line anti tuberculosis injectable drugs—i.e., capreomycin, kanamycin,and amikacin have also been reported [2].

Monitoring the Treatment of TB

Patients should be monitored closely for signs of treatment failure. Monitoring response to treatment is done through regular history taking, physical examination, chest radiograph and laboratory monitoring. The classic symptoms of TB – cough, sputum production, fever and weight loss – generally improve within the first few weeks. Cough and sputum production can persist after sputum conversion in patients with extensive lung damage, but even in those with extensive lung damage improvement is often seen within a month or two of effective treatment. Persistent fever, weight loss or recurrence of any of the classic symptoms of TB should prompt investigation of treatment failure or untreated comorbidities. The recurrence of TB symptoms after sputum conversion may be the first sign of treatment failure. The chest radiograph may appear unchanged in the first few months of treatment or show only slight improvement, especially in patients with chronic pulmonary lesions. Chest radiographs should be taken at least every six months to document progress and to use for comparison if the patient's clinical condition changes.

LITERATURE REVIEW

Richard Appiah and Joseph Kobina Panford *et al.*[1] employs the use of Adaptive Neuro-Fuzzy

Inference System (ANFIS) to provide a better option for malaria diagnosis than the traditional diagnosis method which is characterized by erotic guess work and observation of patients by doctors. Datasets of patients divided into training and checking data were used to train the ANFIS. The results tested after training showed that ANFIS has the ability to diagnose malaria efficiently than the traditional method with very minimal error.

Ajay Kumar Shrivastava and Akash Rajak *et al* [2] describe the designing of intelligent system based on Adaptive Neuro Fuzzy Inference System (ANFIS). The system will detect pulmonary tuberculosis stages based on various input parameters. A system diagram describing the various blocks and flowchart will be designed for the intelligent system. The intelligent system will be rule based and rules are formulated for the diagnosing the various stages of tuberculosis.

Maryam Rezaei Farokhzad and Laya Ebrahimi [3] developed a tendency towards intelligent systems for better diagnosis. Thus, in this paper, we have diagnosed liver sickness using fuzzy logic by obtaining important laboratory parameters. We have used two types of triangular membership function and Gussy membership function per 3 membership function for each input and output and also the design of reference table (search) to construct fuzzy heuristic system and we have managed to compare 243 rules. At the end, we were able to reach to 79/83% of accuracy with the appropriate choice of input parameters, the number and type of membership functions.

Ricky Gogoi and Kandarpa Kumar Sarma [4] development of an information extraction system based on KMC and ANN and an Adaptive Neuro Fuzzy System (ANFIS) based system with the same purpose to achieve enhanced performance as compared to each other. We specially deal with an ANFIS aided by KMC for use with information extraction from satellite images. Experimental results show that such system is effective in dealing with information extraction from river images with forest and sand distribution along its bank.

Prof Swati S Jayade, Prof. D. T. Ingole [5] focus on computing the probability of occurrence of a particular ailment from the medical data by mining it using a unique algorithm which increases accuracy of such diagnosis by combining Neural Networks, Bayesian Classification and Differential Diagnosis all integrated into one single approach. The system uses a Service Oriented Architecture (SOA) wherein the system components of diagnosis, information portal and other miscellaneous services provided are coupled. This algorithm can be used in solving a few common problems that are encountered in automated diagnosis these days, which include diagnosis of multiple diseases showing similar symptoms, diagnosis of a person suffering from multiple diseases, receiving faster and more accurate second opinion, and faster identification of trends present in the medical records.

Ahmed Abdel-Aleem [6] predict the amount of production inventory to trained the FIS model. Then ANFIS is applied and the results from the FIS alone are compared with the ANFIS results. After that these results are compared with the collected data from the cement industry in order to validate the model and determine the superiority of ANFIS over FIS techniques. The comparison between FIS and ANFIS absolute relative errors is illustrated, as the absolute relative errors are calculated based on the industrial data. The error for each method does not exceed 0.0005 which make these methods acceptable to a high degree with the industrial data the validation of the data obtained from the two methods.we can infer that ANFIS is efficient and strongly recommended for solution of inventory control.

N.V. Ramana Murty, M. S. Prasad Babu [7] present artificial neural network and fuzzy logic in pancreatic disease diagnosis based on a set of symptoms. The real procedure of medical diagnosis which usually is employed by physicians was analysed and converted to a machine implementable format. This paper presents an approach to detect the various stages of pancreatic cancer affected patients. Outcome suggests the effectiveness of using neural network over manual detection procedure.

Mustain Billah , Nazrul Islam [8] proposed lung cancer diagnosis system on Adaptive Neuro Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) . This diagnosis system has mainly two steps: Feature extractionreduction and classification. First, lung cancer historical data sets are collected from different hospitals. They are then preprocessed. To reduce the lung cancer features dimensionality, Linear Discriminant Analysis (LDA) is applied. Reduced features are then fed into ANFIS classifier system. Classification accuracy, sensitivity and specificity analysis are performed for performance evaluation of proposed system. Obtained ac-curacy of about 95.4% shows that the proposed intelligent system has a good diagnosis performance and can be used as a promising tool for lung cancer diagnosis.

Tamer UCAR [9] focuses on classification of tuberculosis patients. To make a correct diagnosis of tuberculosis, a medical test must be applied to patient's phlegm. The result of this test is obtained about after a time period of 45 days. The purpose of this study is to develop a data mining solution which makes diagnosis of tuberculosis as accurate as possible and helps

deciding if it is reasonable to start tuberculosis treatment on suspected patients without waiting the exact medical test results or not. It is imperative that, there must be a very accurate classification for this model. Because false positive classified patients will use strong antibiotics for 45 days for nothing and they have to deal with its side affects. And the false negative classified patients' treatment plan will be suspended for 45 days and within this untreated period their disease will get even worse than it is. Therefore, correct prediction of tuberculosis is a very important issue.According to the findings of our study, we concluded that ANFIS is an accurate and reliable method comparing to Bayesian Network, Multilayer Perceptron, Part, Jrip and RSES methods for classification of tuberculosis patients.

X.Y. Djam , G. M. Wajiga *et al* [10] presented for providing decision support platform to malaria researchers, physicians and other healthcare practitioners in malaria endemic regions. The developed FESMM composed of four components which include the Knowledge base, the Fuzzification, the Inference engine and Defuzzification components. The fuzzy inference method employed in this research is the Root Sum Square (RSS). The Root Sum Square of drawing inference was employed to infer the data from the fuzzy rules developed. Triangular membership function was used to show the degree of participation of each input parameter and the defuzzification technique employed in this research is the Centre of Gravity (CoG). The fuzzy expert system was designed based on clinical observations, medical diagnosis and the expert's knowledge. We selected 35 patients with malaria and computed the results that were in the range of predefined limit by the domain experts.

Wahyuni Eka Sari, Oyas Wahyunggoro and Silmi Fauziati [11] focus on the Mamdani, Tsukamoto and Sugeno-types Fuzzy Inference System are applied to assist the tuberculosis diagnosis. The different technique in these three methods is aimed to determine the most appropriate method for such diagnosis. The results show that, of the three types of Fuzzy Inference System, the best model is Sugeno model. Sugeno-type FIS has a better accuracy compared to both Mamdani and Tsukamoto ones at 93%, equivalent to a fault diagnosis in 13 of 180 patients. Here, Mamdani-type FIS is provided the diagnostic accuracy of 89%, equivalent to the fault diagnosis in 20 of 180 patients. On the other hand, Tsukamoto is provided the diagnostic accuracy of 92%, equivalent to fault diagnosis in 15 of 180 patients. Based on the three systems, the most precise output is found in Sugeno-type Fuzzy with a value by 95.1% while for Fuzzy Mamdani and Tsukamoto, it values are 93.4% and 94.5%, respectively. Also, the highest level for the system sensitivity is found in Sugeno with 97.2% in comparison to Tsukamoto FIS by 96.67% and Mamdani at 94.4%.

Krishna Kanth [12] ANFIS strategy is employed to model nonlinear functions, to control one of the most important parameters of the induction machine and predict a chaotic time series, all yielding more effective, faster response or settling times. Also in this paper, we presented the architecture and basic learning process underlying ANFIS (adaptive-network-based fuzzy inference system) which is a fuzzy inference system implemented in the framework of adaptive networks. Soft computing approaches including artificial neural networks and fuzzy inference have been used widely to model expert behavior. Using given input/output data values, the proposed ANFIS can construct mapping based on both human knowledge (in the form of fuzzy if-then rules) and hybrid learning algorithm. In modeling and simulation, the ANFIS strategy is employed to model nonlinear functions, to control one of the most important parameters of the induction machine and predict a chaotic time series, all yielding more effective, faster response or settling times.

Dr. C. Loganathan and K. V. Girija [13] utilizes the training and learning neural networks to find parameters of a fuzzy system based on the symptoms created by the mathematical model. Adaptive learning is the important characteristics of neural networks. Adaptive Neuro Fuzzy Inference System (ANFIS) is used for system identification based on the available data. The main aim of this work is to determine appropriate neural network architecture for training the ANFIS structure inorder to adjust the parameters of learning method from a given set of input and output data. The training algorithms used in this work are Back Propagation, gradient descent learning algorithm and Runge-Kutta Learning Algorithm (RKLM). The experiments are carried out by combining the training algorithms with ANFIS and the training error results are measured for each combination. The results showed that ANFIS combined with RKLM method provides better training error results than other two methods.

Chang Su Lee [14] proposed to address the issue of automatic generation of membership functions and rules with the corresponding subsequent adjustment of them towards more satisfactory system performance. Because one of the most important factors for building high quality of FIS is the generation of the knowledge base of it, which consists of membership functions, fuzzy rules, fuzzy logic operators and other components for fuzzy calculations. The design of FIS comes from either the experience of human experts in the corresponding field of research or input and output data observations collected from operations of systems. Therefore, it is crucial to generate high quality FIS from

a highly reliable design scheme to model the desired system process best.

Jyh-Shing Roger Jang [15] proposed ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input-output data pairs. In the simulation the ANFIS architecture is employed to model nonlinear functions, identify nonlinear components on-linely in a control system, and predict a chaotic time series, all yielding remarkable results. Comparisons with artificial neural networks and earlier work on fuzzy modeling are listed and discussed. Other extensions of the proposed ANFIS and promising applications to automatic control and signal processing are also suggested

Anuradha T. Agrawal and Pankaj S. Ashtankar [16] develop a system which will help in reducing the frequent visits to the clinic and also help in early diagnosis of dangerous diseases. A system must be targeted both for monitoring elderly and for monitoring rehabilitation after hospitalization period and at the same time economically efficient. This paper presents our initial attempts to develop such a system with the help of Adaptive Neuro-Fuzzy Inference System (ANFIS) by adaptive learning mechanism. The MATLAB simulation results indicate that the performance of the ANFIS approach is much important and at the same time easy to implement. The study results are based on the ranges of diagnostic health parameters and the corresponding opinion of the expert. The developed healthcare system can be useful for the elderly and terminally ill patients confined within their homes and at the same time helpful to the pregnant women for their regular checkups without personally visiting to the clinic.

Raafat Fahmy, Hegazy Zaher and Abd Elfattah Kandil [17] outlines the basic differences between the Fuzzy logic techniques, including Mamdani , Sugeno fuzzy inference system models and Adaptive Neuro-Fuzzy Inference System (ANFIS). The main motivation behind this research is to assess which approach provides the best performance for predicting prices of Fund. Due to the importance of performance in Economy, the Mamdani , Sugeno models and ANFIS are compared with the actual values. Fuzzy inference systems (Mamdani , Sugeno and ANFIS fuzzy models) can be used to predict the weekly prices of Fund for the Egyptian Market. The application results indicate that (ANFIS) model is better than that of Mamdani and Sugeno . The results of the three fuzzy inference systems (FIS) are compared.

Atinc YILMAZ and Kursat AYAN [18] determined by studies and research that stress also triggers cancer types. Early precaution is very important for people who have not fallen ill yet with a disease like cancer that has a high mortality rate and expensive treatment. With this study, we expound that the possibility of developing such disease may be decreased and people could take measures against it. For the 3 cancer types selected as pilot work by introducing a fuzzy logic model, the risks for acquiring these cancer types and preliminary diagnosis for the person to remove these risks are presented. After calculating the risk outcome, the effect of stress on cancer is discussed and determined. Within the study, a fuzzy logic technique that can easily be adapted to other industry studies, as well, is applied to the health industry and effective software for application is developed. Due to this type of study, people will have the chance to take measures against developing cancer and the rate of suffering from cancer may be decreased. Furthermore, the performance status of the new technique is revealed by calculating performance measurements by the outcomes of the models developed by the new type of fuzzy logic technique for 3 cancer types selected as a pilot in the Mamdani type of fuzzy logic model.

Asha.T , Dr. S. Natarajan and Dr. K.N.B. Murthy [19] apply ensemble classification techniques, Adaboost, Bagging and Random forest for classifying tuberculosis (TB). The data is obtained from the state hospital which mainly includes twelve preliminary symptoms (attributes).The data is classified into two categories namely pulmonary tuberculosis (PTB) and retroviral PTB i.e. TB along with AIDS. Evaluation measures such as sensitivity, specificity and accuracy are used for comparison. Random forest is found to be weak with 93% accuracy against 97% that of Bagging and 96% of Adaboost.

Tamer Uçar , Adem Karahoca [20] proposed the use of Sugeno-type “adaptive-network-based fuzzy inference system” (ANFIS) to predict the existence of mycobacterium tuberculosis. 667 different patient records which are obtained from a clinic are used in the entire process of this research. Each of the patient records consist of 30 separate input parameters. ANFIS model is generated by using 500 of those records. We also implemented a multilayer perceptron and PART model using the same data set.

The ANFIS model classifies the instances with an RMSE of 18% whereas Multilayer Perceptron does the same classification with an RMSE of % 19 and PART algorithm with an RMSE of % 20. ANFIS is an accurate and reliable method when compared with Multilayer Perceptron and PART algorithms for classification of tuberculosis patients. This study has contribution on forecasting patients before the medical tests.

K. Soundararajan, Dr. S. Sureshkumar and C. Anusuya [21] main focus is to development of the

system on the architecture and algorithm used to find the probable class of tuberculosis a patient may have. A Rule–based Fuzzy Diagnostics Decision Support System is used to assign class labels for tuberculosis and fuzzy logic technique is used for class assignment process, Fuzzy rule sets are prepared by doctors. Tuberculosis symptoms and class details are updated in rule based system .Learning and testing operations are performed by this process. The system is designed to detect class of tuberculosis and these fuzzy rules are updated using rule mining techniques. Based on this method that generates classes of tuberculosis suits the needs of pulmonary physicians and reduce the time consumed in generating diagnosis.

Rusdah, Edi Winarko [22] classified by variables, data preprocessing techniques and methods used for tuberculosis diagnosis. From those selected literatures that have been reviewed, we conclude that the most frequently used variables are sweating at night, more than 3 weeks of cough, fever, weight loss, age, and chest pain respectively. Support Vector Machine gave the highest accuracy 98,7%, followed by Bagging 98,4% and RandomForest 98,3% compared to other methods. In addition, several experiments have shown that ANFIS is the most accurate method in diagnosing tuberculosis.

K. R. Lakshmi , M. Veera Krishna and S. Prem Kumar [23] A main goal medical data mining algorithm is to get best algorithms that describe given data from multiple aspects. The algorithms are very necessary for intend an automatic classification tools. The PLS-DA was the best one among ten (five criteria are satisfied). PLS-DA classifier is suggested for Tuberculosis disease based classification to get better results with accuracy and performance.

Shakshi Garg, Navpreet Rupal [24] objective of this research work is to create a data mining way out that makes identification of TB as exact as possible. In our proposed framework we have used various techniques such as centroid selection based clustering algorithm would be used to enhance the clustering scheme, PCA for feature extraction, genetic algorithm for feature optimization and neural network for training and testing purpose. In the end, results are being evaluated after classification and testing on the basis of performance parameter such as accuracy, recall, precision, false acceptance ratio, and false rejection ratio.

Rupali Zakhmi [25] focused on medical diagnosis of tuberculosis rigorous. There are assorted parameters such as Cough, Chest Pain, Night Sweats, Age, Weight Loss, Gender and Fever, Coughing up Blood, No Appetite which are used for predicting tuberculosis. Both Genetic algorithm and Neural network backwash better than other techniques. Tuberculosis disease forecasting is accomplished by soft computing technique. Genetic algorithm offers best fitness value, disembroil optimization problems whereas Neural Network takes parameters as input and also utilize genetic operators to train the neural network and spawn an output for presaging tuberculosis disease. This research outlines the main review and technical papers on tuberculosis detection that are implemented using multifarious data mining techniques. Review of papers surmises that soft computing technique acquires the highest accuracy.

Sedentary pursuits, such as watching television and using the computer, are believed to be an important environmental factor contributing to the fact that 25% of children in the United States are overweight or obese [30].

Jingming Liu, Wei Wang, Jing Xu, Mengqiu Gao, Chuanyou Li [26] proposed on both MDR-TB and snMDR-TB simultaneously. Defined the concept of snMDR-TB patients as SN-PTB patients whose clinical profiles are similar to those of MDR-TB patients, and have the potential possibility to become a MDR-TB patient. The basic issues about snMDR-TB are how to determine whether a patient is snMDR-TB and how many sn MDR-TB cases there are in the real world. In this article we apply statistical learning methods to explore these problems.

Negar Ziasabounchi and Iman Askerzade [27] develop a method of classifying for heart disease degree of patient based characteristic data using adaptive neuro fuzzy inference system. The data were obtained from the University of California at Irvine (UCI) machine learning repository. Seven variables are used as input of prediction model. To test the ability of the trained anfis models to recognize heart disease diagnosis, we used k-fold cross validation method. The experimental results demonstrate that the model successfully forecasts the patient's heart disease degree with an accuracy rate of 92.30%.

Niranjan Pramanik and Rabindra Kumar Panda [28] estimate the flow at the downstream stretch of a river using flow data for upstream locations using ANFIS. Three feed-forward back-propagation training algorithms were used to train the models. Standard performance indices, such as correlation coefficient, index of agreement, root mean square error, modelling efficiency and percentage deviation in peak flow, were used to compare the performance of the models, as well as the training techniques. The results revealed that the neural network with conjugate gradient algorithm performs better than Levenberg-Marquardt and gradient descent algorithms. The model which considers as input the reservoir release up to three antecedent time steps

produced the best results. It was found that barrage outflow could be better estimated by the ANFIS than by the ANN technique.

Shaikh Abdul Hannan, V.D. Bhagile *et al.* [29] develop a expert system for diagnosing of heart disease using support vector machine and feedforward backpropagation technique. Now a days neural network are being used successfully in an increasing number of application areas. This work includes the detailed information about patient and preprocessing was done. The Support Vector Machine (SVM) and feedforward Backpropagation technique have been applied over the data for the expert system. To make the system more authentic and reliable out of 300 patients 250 patients were used for training set and 50 for evaluation process. In conclusion, we have used two neural network techniques but we are getting just 50% to 60% output i.e. not reliable for the patient. This expert system data can also be applied to improve the accuracy the medicine using some other neural network techniques.

ADAPTIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS)

During the late 1980s, the number of researchers and engineers interested in neural networks (NNs) and fuzzy logic (FL) increased, dramatically introducing the NN and FL technologies into several application fields. Both technologies are widely used and are considered fundamental engineering technologies. Within several years, NN and FS fusing technologies were already being used in commercial products and industrial systems. Today these techniques are very popular in biomedical field like medical diagnosis.

Adaptive Neuro Fuzzy Inference System (ANFIS) is a kind of hybrid of neural network and fuzzy logic and is based on fuzzy inference system. In ANFIS, we combine both the learning capabilities of a neural network and reasoning capabilities of fuzzy logic in order to give enhanced prediction capabilities [2]. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions. Hence, ANFIS is considered to be universal approximator.

The ANFIS model is very suitable and can generate excellent classification results provided that the right type and number of Membership Functions (MFs) are used in the classification task [2]. In the classification two different classification techniques are employed: an artificial neural network-based classifier and a hybrid ANFIS classifier. A neural classifier can learn from data, but the output does not lead itself naturally to interpretation. An ANFIS classifier is based on a three-layer feed-forward neural network and combines the merits of both neural and fuzzy classifiers while overcoming their drawbacks and limitations. The developed Adaptive Neuro Fuzzy Inference System (ANFIS) classifier exhibits high levels of accuracy, consistency and reliability, with acceptably low computational time and is a promising new development in the field of diagnosis of tuberculosis.

METHODOLOGY

K-Means is one of the unsupervised learning algorithms that solve the widely-known clustering problem. The purpose of this algorithm is to define K centers for each cluster. These centers should be placed in a cunning way because of different location causes different result. so the better choice is to place them as much as possible for away from each other. Here we take each point belongs to a given dataset and associate it to the nearest center until no points are in pending. The first step is completed and an early group age is done. At this point we need to re-calculated k new centroids as a center of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we obtained the k centers of their location by using the step by step process until no more changes are occur.

ANFIS

ANFIS is a neural-fuzzy system which contains both neural networks and fuzzy systems. A fuzzy-logic system can be described as a non-linear mapping from the input space to the output space. This mapping is done by converting the inputs from numerical domain to fuzzy domain. To convert the inputs, firstly fuzzy sets and fuzzifiers are used. After that process, fuzzy rules and fuzzy inference engine is applied to fuzzy domain . The obtained result is then transformed back to arithmetical domain by using defuzzifiers. Triangular functions are used for fuzzy sets and linear functions are used for rule outputs on ANFIS method. The standard deviation, mean of the membership functions and the coefficients of the output constant functions are used as network parameters of the system.

The summation of outputs is calculated at the last node of the system. The last node is the rightmost node of a network. In Sugeno fuzzy model, fuzzy if-then rules are used. The following is a typical fuzzy rule for a Sugeno type fuzzy system:

If x is A and y is B then x = f(x, y)

In this rule, A and B are fuzzy sets in anterior. The crisp function in the resulting is z=f(x, y). This function mostly represents a polynomial. But exceptionally, it can be another kind of function which can properly fit the output of the system inside of the

fuzzy region that is characterized by the anterior of the fuzzy rule. We use first-order Sugeno fuzzy model for cases which are having f(x, y) as a first-order polynomial.

Let's scrutinize a first-order Sugeno fuzzy inference system having two rules:

Rule 1: If X is A1 and Y is B1, then f1 = p1x + q1y + r1

Rule 2: If X is A2 and Y is B2, then f2 = p2x + q2y + r1

In the following Figure 1, the fuzzy reasoning system is illustrated in a shortened form. In order to bypass excessive computational complexity in the process of defuzzification, only weighted averages are used.

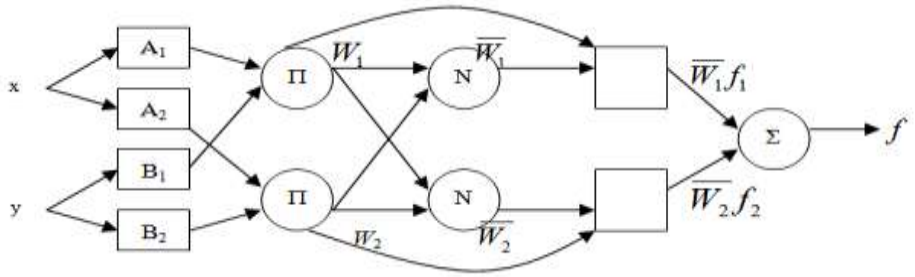

Fig-2 : ANFIS Architecture

ANFIS System Training Process

The ANFIS system training methodology is summarized in Figure 3. The process begins by obtaining a training data set (input/output data pairs) and testing data sets. The training data is a set of input and output vectors. Two vectors are used to train the ANFIS system: the input vector and the output vector. The training data set is used to find the premise parameters for the membership functions. A threshold value for the error between the actual and desired output is determined. The consequent parameters are found using the least squares method. If this error is larger than the threshold value, then the premise parameters are updated using the gradient decent method. The process is terminated when the error becomes less than the threshold value [16].

ANFIS training learning rules use hybrid learning, combining the gradient descent and the least squares method. The aim of using ANFIS for health monitoring is to achieve the best performance possible. ANFIS training begins by creating a set of suitable training data in order to be able to train the Neuro-Fuzzy system. The data set used as the input to the anfis function must be in a matrix form, where the last column in the matrix is the output, and the matrix contains as many columns as needed to represent the inputs to the system. The rows represent all the existing data situations. Creation of the membership functions is dependent on the system designer. The designer may create the parameters of the membership functions if they have knowledge of the expected shapes, or they can use the command genfis1 from MATLAB to help in the creation of the initial set of membership functions. This work uses the genfis1 command to create the membership functions. Once the initial membership functions are created, system training begins. When the training process is finished the final membership functions and training error from the training data set are produced. After the system training is complete, ANFIS provides a method to study and evaluate the system performance by using the evalfis function. Once the ANFIS is trained, we can test the system against different sets of data values.

ANFIS for Monitoring Pulmonary Tuberculosis

ANFIS is selected to solve the problem of remote Pulmonary Tuberculosis monitoring. The steps required to apply ANFIS to modeling are: define input and output values; define fuzzy sets for input values; define fuzzy rules; and create and train the Neural Network. To implement and test the proposed architecture, a development tool is required. MATLAB Fuzzy Logic Toolbox (FLT) from MathWorks was selected as the development tool. This tool provides an environment to build and evaluate fuzzy systems using a graphical user interface. It consists of a FIS editor, the rule editor, a membership function editor, the fuzzy inference viewer, and the output surface viewer. The FIS editor displays general information about a fuzzy inference system. The membership function editor is the tool that displays and edits the membership functions associated with all input and output variables. The rule editor allows the user to construct the rule statements automatically, by clicking on and selecting one item in each input variable box, one item in each output box, and one connection item. The rule viewer allows users to interpret the entire fuzzy inference process at once. The ANFIS editor GUI menu bar can be used to load a FIS training initialization, save the trained FIS, and open a new Sugeno system to 1interpret the trained FIS model.

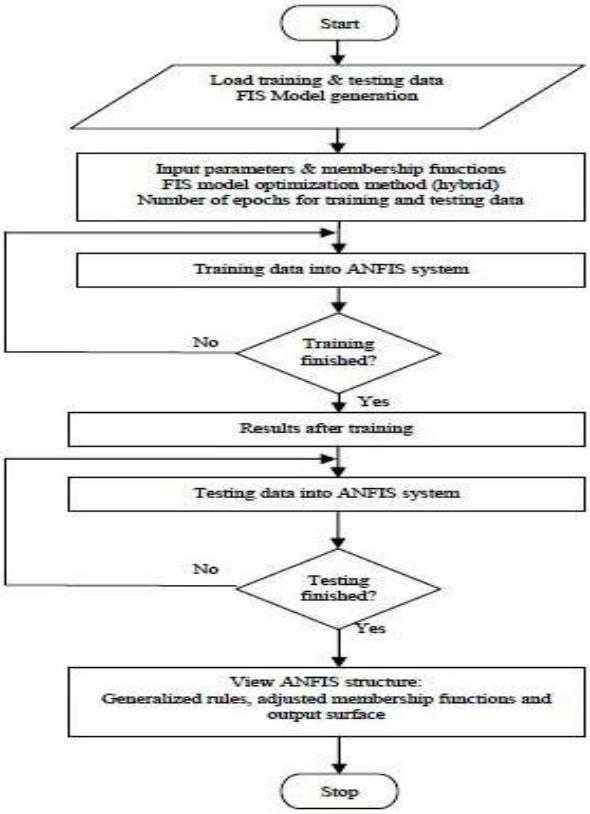

Fig-3: Steps for Monitoring Pulmonary Tuberculosis

IMPLEMENTATION OF ANFIS

Step 1:

Totally collected tuberculosis dataset is divided into two groups using K-Means Clustering algorithm. Cluster-1 consist 73% of Pulmonary tuberculosis dataset and Cluster2 consist 44% of Extra pulmonary tuberculosis dataset. The maximum number of dataset is Pulmonary tuberculosis dataset, which is consider as Cluster-1 is taken for furthermore implementation of ANFIS based medical diagnosis.

Step 2:

The analysis process of Fuzzy Inference System Sugeno is done by MATLAB software. thus we import Sugeno Fuzzy Inference System to generate the crisp output for an input fuzzy , which use weighted average to calculate the crisp output.

Step 3:

Fuzzification

Conducting fuzzification is made of the membership function of each input variable. From the membership function, the crisp value is converted into fuzzy values by means of a fuzzification technique. Input variable membership function plot is shown as below.

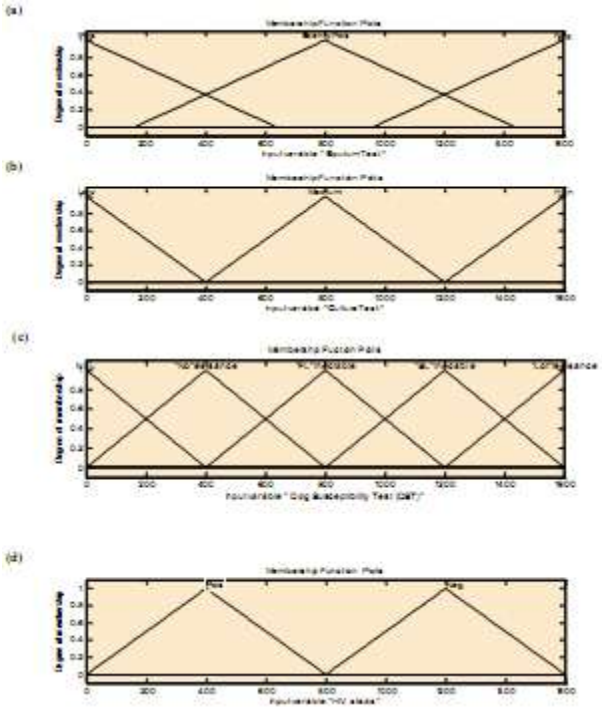

Fig-4: The Membership Function of Variable inputs; (a) Sputum ; (b) culture; (c) Drug Susceptibility Test (DST) ; (d) HIV status

Step 4:
Generate Fuzzy Rules

Table 1:

Rule No	Antecedent	Consequent
1	If Drug Susceptibility Test is First Line(FL)Injectable AND Culture is Low	MDR-TB
2	If Drug Susceptibility Test is Second Line(FL)Injectable AND Culture is Medium	XDR –TB
3	If Drug Susceptibility Test is Lot Resistance AND Culture is High	XXDR –TB
4	If HIV Test is Positive AND Sputum Test is Positive	HIV –TB
5	If Drug Susceptibility Test is First Line(FL)Injectable AND Culture is Medium	MDR –TB
6	If Drug Susceptibility Test is Second Line(FL)Injectable AND Culture is Low	XDR –TB
7	If Drug Susceptibility Test is First Line(FL)Injectable AND Culture is High	MDR- TB
8	If HIV Test is Positive AND Sputum Test is Scanty Positive	HIV –TB
9	If Drug Susceptibility Test is Second Line(FL)Injectable AND Culture is High	XDR –TB

Based on fuzzy rules, a consequent can be classified into Four outcomes: Multi Drug Resistant tuberculosis (MDR-TB), Extensively Drug Resistant (XDR-TB), Extremely Drug Resistant (XXDR-TB) and HIV-TB.

Step 5:
Modeling with the ANFIS Editor

- First dataset that contains desired input/output data pairs of the target system to be modeled were collected. The datasets are now divided into training and checking datasets.
- The training and checking datasets are saved in excel files (or) Save files in Notepad

- The training and checking datasets are imported individually into Matlab workspace using the command ***uiimport*** in the Matlab command area.
- The command **anfisedit** was typed in the Matlab command area to display ANFIS editor dialogue box.
- In the load data section of the ANFIS editor, training and checking data were loaded by selecting appropriate radio buttons and then clicking **Load Data**. The loaded data is plotted on the plot region.
- FIS model was generated by clicking on grid partition in the **Generate FIS** section of the ANFIS editor.
- FIS model structure was viewed once an initial FIS has been generated or loaded by clicking the **Structure** button.
- The FIS model parameter hybrid optimization method: which is a mixture of back propagation and least squares method was chosen in the **Train FIS** section of the ANFIS editor. The error tolerance and the training epochs number were also chosen in this section.
- FIS model was trained by clicking the **Train now** button. This training adjusted the membership function parameters and plotted the training data error plots in the plot region.
- The **Test button** in the "Test FIS" portion of the ANFIS editor was clicked to view the testing plot against the training dataset.

Table 2: Portion of the dataset for training in ANFIS

Patient ID	DST	CULTURE	HIV	SPUTUM	Level of TB
T001	**500**	**300**	**0**	**0**	**1**
T002	**600**	**300**	**0**	**0**	**1**
T003	**900**	**500**	**0**	**0**	**2**
T004	**1600**	**1600**	**0**	**0**	**3**
T005	**0**	**0**	**300**	**300**	**4**
T006	**0**	**0**	**600**	**1400**	**4**
T007	**1500**	**500**	**0**	**0**	**2**
T008	**1300**	**1300**	**0**	**0**	**3**
T009	**1500**	**1100**	**0**	**0**	**2**
T010	**0**	**0**	**700**	**600**	**4**

Table 2 displays the snapshot of the portion of dataset used in training the ANFIS. Sputum, culture, Drug Susceptibility Test , HIV status were used as input variables Whereas Level of Intensity was used as output variable.

Step 6:
COMPARING RESULTS OF THE EXPERIMENTATION

The purpose of the comparison is to find out how accurate or close the results of ANFIS are close to the expected degree or level of intensity of the Tuberculosis condition. Table 3 shows the columns for ID of some selected patients through random picking, expected degree of Tubercuosis condition, ANFIS result and ANFIS Diagnosis.

Table 3: Comparison of results of ANFIS and expected level of Tuberculosis condition

Patients ID	Expected Result	ANFIS Result	ANFIS Diagnosis
T012	1	1.12	MDR-TB
T021	3	3.3	XXDR-TB
T014	4	4	HIV-TB
T028	2	1.92	XDR-TB
T019	1	0.937	MDR-TB
T013	1	1.7	MDR-TB
T028	2	1.92	XDR-TB
T023	1	1.03	MDR-TB
T020	2	1.82	XDR-TB

DISCUSSION OF RESULTS

Table 3 shows the ANFIS results of some selected patients. Also given the values for the linguistic input variables, degree or intensity of the Tuberculosis condition were assigned to them as output to be modeled in ANFIS. The intensities MDR-TB, XDR-TB, XXDRTB and HIV-TB were coded into figures 1, 2 and 3 respectively. The result after training in table 2 shows very good results. Though there are some differences between the expected tuberculosis intensity and the ANFIS, it can be confidently said that the errors are very minimal. For example the result of

patient T019 after the trained ANFIS is 0.937 which is undoubtedly close to 1. Again patient T014 had result 4 after the trained ANFIS which is also very exact to the expected result which is 4. Patient T014 had 100% accuracy in diagnosis since 4 was obtained after ANFIS training equaling the expected result which is 4. The researchers realized from the experiment conducted on the sampled patients that their results of level or intensity of pulmonary tuberculosis condition were very close to the expected degree of tuberculosis condition used as output to train the dataset.

CONCLUSION AND FUTURE ENHANCEMENT

Data mining techniques are used to extract useful information and knowledge from the database. Classification is a data mining technique which explores data from the training set and builds a classifier model, based on these data, which can be used for performing predictions. The genetic algorithms are adaptive techniques that can be successfully used in solving complex search and optimization problems. A database can be viewed as a very large search space and a genetic algorithm can be viewed as a search strategy in a database. The classification rules that are searched are the ones that will constitute the classifier model. In this paper, different classifiers are studied and the experiments are conducted to find the best classifier for predicting the patient of heart disease. We propose an approach to predict the heart diseases using data mining techniques. The empirical results show that we can produce short but accurate prediction list for the heart patients by applying the predictive models to the records of incoming new patients. This prediction model helps the doctors in efficient heart disease diagnosis process with fewer attributes. Heart disease is the most common contributor of mortality in India. This study will also work to identify those patients who needed special attention.

In order to improve the performance of mining the genetic and knn algorithm. Follow the steps :

i. predict early in the heart disease, it will improve the Classification accuracy for many datasets and especially in our country.

ii. Preventive Strategies to reduce risk factors are essential and to reduce the alarmingly increasing burden of heart disease in our population.

iii. and Identification of major risk factors and developing decision support system, and effective control measures and health education programs will decline the heart disease mortality.

Data Set for analysis may contain number of attributes, many of which may be irrelevant or redundant features. These features are eliminated to improve the accuracy and to reduce the time complexity of the K-NN classifier algorithm. Here we are collected the sample Data, then using the Genetic algorithm and K-NN classifier algorithm.

REFERENCES

1. Oswal A, Shetty V, Badshah M, Pitre R, Vashi M. A survey on disease diagonsis algorithms. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). 2014; 3(11).
2. Ashari A, Paryudi I, Tjoa AM. Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool (IJACSA). International Journal of Advanced Computer Science and Applications. 2013;4(11).
3. Jabbar AM, Deekshaulu BL. "Heart disease classification using nearest neighbor classification with feature subset selection. Anale. Seria Informatică.2013; XI(1).
4. Chaudhari AA, Akarte SP. Fuzzy & Datamining based Disease PredictionUsing K-NN Algorithm. International Journal of Innovations in Engineering and Technology (IJIET).
5. Ahmed A, Hannan SA. Data Mining Techniques to Find Out Heart Diseases: An Overview. International Journal of Innovative Technology and Exploring Engineering (IJITEE). 2012;1(4).
6. Behrouz Minaei, William F. "Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System", Genetic Algorithms Research and Applications Group (GARAGe).
7. Bahrami B, Shirvani MH. Prediction and Diagnosis of Heart Disease by Data Mining Techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST). 2015; 2(2).
8. Carvalho DR, Freitas AA. A Hybrid Decision Tree/Genetic Algorithm Method for Data Mining. proceedings of the Conference on Emerging Artificial Inteelligence Applications in Computer Engineering, 2007.
9. Fix E, Hodges J. Dicriminatory analysis ,non parametric discrimination:consistency properties. Technical report 4,USA,School of aviation medicine Randolph field texas, 1951.
10. Goldberg DE. Genetic algorithm in search optimization and machine learning. Addison Wesley,1989.
11. Masethe HD Masethe MA. Prediction of Heart Disease using Classification Algorithms. Proceedings of the World Congress on Engineering and Computer Science. 2014; II, San Francisco, USA.
12. Jabbar M A, Deekshatulu BL, Chandra P. Heart disease prediction system using associative classification and genetic algorithm. pp183-192 Elsevier, 2012.
13. Jain R, Mazumdar J. A genetic algorithm based nearest neighbor classification to breast cancer

diagnosis. Australasian Physics & Engineering Sciences in Medicine March 2003, 26:6.
14. Dotri J, Renugadevi T. Analysis of various datamining technique. Indian journal of science and technology. 2012; 5(35).
15. Jayavani K, Kadhar Nawaz GM. Optimal Data Prediction and Classification Applicable for Intelligent Heart Disease Diagnosis System. International Journal of Computational Intelligence and Informatics2015; 5(2).
16. Soni J, Ansari U. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. International Journal of Computer Applications (0975 – 8887). 2011;17(18).
17. Kaliya Meiyar V, Shanmugasundaram D. The Comparative Study for Diagnosing Heart Disease Using KNN and Naïve Bayes. International Journal of Advance Research in Computer Science and Management Studies,2015; 3(8).
18. Kavitha R, Kannan E. A Framework for Heart Disease Prediction Using K nearest Neighbor Algorithm. Research Journal of Applied Sciences, Engineering and Technology. 2015; 10(1): 10-13.
19. Wisaeng K. "Predict the Diagnosis of Heart Disease Using Feature Selection and k-Nearest Neighbor Algorithm. Applied Mathematical Sciences. 2014;8(83): 4103 – 4113.
20. Lashari SA, Ibrahim R. Comparative analysis of data mining techniques for medical data classification. Proceedings of the 4th International Conference on Computing and Informatics, 2013;28-30.
21. Pradhan M, Sahu RK. Predict the onset of diabetes disease using Artificial Neural Network (ANN). International Journal of Computer Science & Emerging Technologies (E-ISSN: 2044-6004). 2011; 303(2).
22. Metkari M, Pradhan M. "Improve the Classification Accuracy of the Heart Disease Data Using Discretization. International Journal of Innovative Research in Advanced Engineering (IJIRAE). 2015; ISSN: 2349-2163.
23. Manimekalai K. Prediction of Heart Diseases using Data Mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization). 2016; 4(2).
24. Michael W, Berry. Lecture notes in data mining. World Scientific, 2006.
25. Mohanraj E. "Heart Disease Prediction using K Nearest Neighbour and K Means Clustering. International Journal of Advanced Engineering Research and Science (IJAERS). 2016; 3(2).
26. Abdar M, Niakan Kalhori SR, Sutikno T, Ibnu Subroto IM, Arji G. "Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. International Journal of Electrical and Computer Engineering. 2015; 5(6): 1569~1576.
27. Dey M, Rautaray SS. Study and Analysis of Data mining Algorithms for Healthcare Decision Support System".International journal of Applied Engineering research. 2012;7(2).
28. Waghulde NP, Patil NP. Approach for Heart Disease Prediction. International Journal of Advanced Computer Research. 2014;4(3)16.
29. Vandana NB. Survey on nearest neighbor techniques. IJCSIS. 2010; 80.
30. Ashadevi B, MuthamilSelvi P, Menaka R. "The impact of computer use on Adolescent's Activities and Development", International Journal of Engineering Science Invention Research & Development. 2017;IV(I).