Neural Comput & Applic (2013) 23:471–483
DOI 10.1007/s00521-012-0942-1

ORIGINAL ARTICLE

Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets

Tamer Uçar · Adem Karahoca · Dilek Karahoca

Received: 30 December 2011 / Accepted: 12 April 2012 / Published online: 29 April 2012

Abstract A correct diagnosis of tuberculosis disease can be only stated by applying a medical test to patient's phlegm. The result of this test is obtained after a time period of about 45 days. The purpose of this study is to develop a data mining solution that makes diagnosis of tuberculosis as accurate as possible and helps deciding whether it is reasonable to start tuberculosis treatment on suspected patients without waiting for the exact medical test results. We proposed the use of Sugeno-type "adaptive-network-based fuzzy inference system" (ANFIS) to predict the existence of mycobacterium tuberculosis. Data set collected from 503 different patient records which are obtained from a private health clinic (consent of physicians and patients). Patient record has 30 different attributes which covers demographical and medical test data. ANFIS model was generated by using 250 records. Also, rough set method was implemented by using the same data set. The ANFIS model classifies the instances with correctness of 97 %, whereas rough set algorithm does the same classification with correctness of 92 %. This study has a contribution on forecasting patients before the medical tests.

Keywords Tuberculosis disease diagnosis · ANFIS · Roughs sets · Data mining

T. Uçar · A. Karahoca (✉) · D. Karahoca
Software Engineering Department, Bahcesehir University,
Istanbul, Turkey
e-mail: akarahoca@bahcesehir.edu.tr

T. Uçar
e-mail: tamer.ucar@bachesehir.edu.tr

D. Karahoca
e-mail: dilek.karahoca@bahcesehir.edu.tr

1 Introduction

Tuberculosis (TB) was believed to be almost under control, but it has once again become a serious worldwide problem. Tuberculosis disease is caused by a bacterium which is called as mycobacterium tuberculosis. This disease can spread among humans, and patients who suffer from tuberculosis might die unless they get the right treatment. This microorganism widely exists on humans, cattle, sheep and birds. All of the organs in the body can be affected by tuberculosis. But most of the tuberculosis cases are occurring in lungs [1].

Tuberculosis disease occurs under different manifestations on adults and children. When the first encounter happens with bacillus, which mostly happens on the childhood phase of a person, lymphatic glands that are located at the entry point of the lungs are picked by this microorganism for the first rooting point on the body. As a result of this event, those glands enlarge (hilar lymphadenopathy) and this called as primary tuberculosis. The adult type (secondary) tuberculosis is different from this scenario. In those cases, the person's lung is contaminated with the microorganism before. If the immune system is strong enough, microorganism cannot cause any sickness but can keep itself alive. When the immune system of the person weakens for a reason, microorganism gets activated and begins to create sickness. Prostration, long-term sicknesses, insomnia, tobacco and alcohol abuse, drug addiction, having an irregular life, malnutrition and stress are some factors which are responsible for weakening the immune system and providing a suitable basis for illness to occur. Unlike primary tuberculosis, lesions are spread to lung parenchyma tissue in secondary tuberculosis cases. Cavities (holes) which may cause lung tissue to bleed can also be seen on advanced phases of the illness [2].

Lung tuberculosis can be seen on very wide age range. From new born babies to old people, everybody can be affected by this disease. Symptoms are cough, fatigue, exhaustion, anorexia, night sweating, fever (which not exceeds 37.5 °C), cavities and haemoptysis on advanced cases [3].

World Health Organization's Direct Observation of Therapy is the internationally accepted standard for control of tuberculosis (http://www.who.int/tb/en/). To make an exact diagnosis, existence of microorganism in phlegm must be proven. But some other microorganisms can also be flagged as mycobacterium tuberculosis under microscope observation. In order to avoid this problem, a special culture medium is prepared where only bacteria of mycobacterium tuberculosis can reproduce. The phlegm sample which is obtained from a patient is planted to this medium and kept for 45 days at body temperature. At the end of this time period, the culture medium is checked for any reproduction sign of the bacteria.

In order to cure tuberculosis, 4–5 different major anti-tuberculosis antibiotics are used for 6–12 months. Some cases may heal without any treatment plan if the immune system is strong enough. After full recovery, lung wounds which are caused by tuberculosis disease still exist as calcific tissue. Unfortunately, cases which are not treated may result by death of patient [2].

According to the report of The Turkish Federation of Anti-Tuberculosis Associations, in 2006, 20,526 patients were suffered from tuberculosis in Turkey. Among, 89.4 % of these patients were treated successfully, and 4.7 % of the patients quit the treatment program. 2.9 % of the patients were having an ongoing treatment program, and 3.0 % of the whole tuberculosis patients died by this disease [4].

As it is stated by Pena et al. [5], data mining can be defined as the process of generating valuable and intelligible knowledge which was not known previously. From this viewpoint, Bakar and Febriyani applied rough neural networks (RNN) for classifying the tuberculosis types. Data set has 233 records, which has 14 attributes, firstly reduced as a result of preprocessing of data. The decisive data set contains 8 attributes, which are gender, age, weight, fevers, night sweats, (cough >3 weeks), blood phlegm and sputum test. 70 % of the data set (131 instances) is used for training, and 30 % (56 instances) is used for testing. Discretization is applied on the numeric and continuous attributes using rough set application. After then, neural network is applied for training and testing the data. Classification accuracy of Tuberculosis is (92.29 %) with RNN model [6].

Sanchez et al. [7] implemented data mining technique to classify TB-related handicaps to determine patients' sickness. This study was classifying tuberculosis diagnostic categories based on given variables. Records of 1,655 patients having 56 attributes are used as raw data set. Those 56 attributes are reduced into 5 attributes which are antecedents, bacteriology results, age category, pulmonary tuberculosis and extra pulmonary tuberculosis. Exhaustive Chi-squared automatic interaction detector (CHAID) is selected for generating decision trees for classes [7].

A recent study is made on diagnosis of chest diseases (tuberculosis, COPD-chronic obstructive pulmonary disease-, pneumonia, asthma, lung cancers, normal) with a data set that have 357 samples in order to classify patients into six classes. Multilayer, probabilistic, learning vector quantization and generalized regression neural networks were applied on the data set. Classification accuracy of Tuberculosis is 95.08 % with the multilayer neural network model [8].

Another research is made on a computer aided diagnosis system. The aim of the system was performing diagnosis on medical images. Multiple classification ripple down rule was performed. The researchers tried to show that it is possible to integrate both detection and diagnosis systems as a new computer aided diagnosis architecture [9].

Rough sets could be a useful methodology to deal with data with uncertainty. In the study of Yang and Wu [10], rough sets applied to identify the set of significant symptoms causing diseases and to induce decision rules using the data of a Taiwan's otolaryngology clinic. Rough set method was selected as artificial learning model, and 657 records with 12 attributes were used for training and testing the system. Classification accuracy of all attributes is 56.03 %.

A multilayer perceptron-based system was proposed by the authors to diagnosing heart disease. System accepts 40 parameters as input, and 352 medical records were used to train and test the system. Authors concluded that proposed system generates accuracy higher than (90 %) [11].

Diagnosis of diabetes by using adaptive neuro fuzzy inference systems (ANFIS) is another application of data mining. Karahoca et al. [12] focused on the fact that determining the high risks of diabetes is the best method for permeating it. According to this fact, the aim of this research is estimating diabetes risk depending on some variables such as age, total cholesterol, gender or shape of the body by using ANFIS. The data set has 390 records, and each record has four variables. The 300 of those records are used for training, and 90 are used for checking the model. Classification accuracy is 87 % for ANFIS to detect diabetes disease.

Another data mining approach for a biomedical topic is the classification of cardiac beat using a fuzzy inference system. For training and testing data sets, MIT Arrhythmia Database and in vivo records from cardiac voluntary patients were used. The point of Monzon and Pisarello's study is identifying and classifying normal versus premature ventricular contractions (PVC). Data set has 34 records and contains 4,917 PVCs and 55,508 normal beats. 2,027 beat data that have 520 PVCs are used for training ANFIS [13].

ANFIS also used to diagnosis Alzheimer disease under the topic as Analysis of MEG. Background activity in Alzheimer's disease is detected by using nonlinear methods and ANFIS. Gomez et al. [14] intend to analyse magneto encephalogram background action on patients using sample entropy and Lempel–Ziv complexity.

Shlomi et al. [15] studied for predicting metabolic biomarkers of human inborn errors of metabolism. The motivation is to get the genome-scale network model of human metabolism. In the light of this event, researchers offer a novel computational approach for systematically predicting metabolic biomarkers in stoichiometric metabolic models.

It was proposed the use of adaptive-network-based fuzzy inference system (ANFIS) to predict the level of cyclosporine A in blood of renal transplantation patients. The aims of Gören et al. [16] are predicting the results of the therapeutic drug monitoring (TDM) process with the help of ANFIS. Data were collected from 138 patients, each containing 20 input parameters. Both Takagi and Sugeno-type ANFIS is used to predict the concentration of Cyclosporine A in blood samples.

It is imperative that there must be high sensitivity and specificity results for data mining models. False positive classified patients will use strong antibiotics for 45 days for nothing and they have to deal with its side effects. 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 an extremely serious issue to predict at the first phase of the disease. For these reasons, in this study, ANFIS [17] and rough sets [18] are considered and benchmarked to obtain most sensitive, and correct rule based classification model.

2 Materials and methods

In this section, preparing tuberculosis data set and implementation of the ANFIS and rough sets are considered for diagnosing the disease of the TB patients. The ANFIS and rough set models were constructed with 20 inputs.

2.1 Preparing tuberculosis data set

In order to obtain the best prediction model for the tuberculosis disease, data set was collected from Private Health Clinic in Istanbul. Patient examination reports were not in electronic format; therefore, 503 patients' electronic health records inserted into a database system. The data set covers examination reports of patients were examined between 01.01.2000 and 31.12.2009. As listed in Table 1, data set was segmented into five different classes based on the value of the output. Output represents the possibility of the existence of the bacteria.

To make an exact diagnosis for patients from classes 0.25, 0.50 and 0.75, existence of microorganism in patient's phlegm must be proven by applying a medical examination. Each patient record consists of 30 different variables. All of the variables are based on World Health Organization's standard of Direct Observation of Therapy. The full list of the variables is listed in Table 2.

2.1.1 Domain values of input variables

Attributes of the data set can be categorized into three groups as follows: (1) demographics data and clinical findings, (2) medical laboratory findings and (3) radiological findings.

In the first group, the *gender* parameter indicates whether the patient is male or female. *Age group* indicates the age group that patient belongs to. All ages are grouped into seven different classes. These classes are "18–24", "25–32", "33–40", "41–45", "46–51", "52–57" and "58+". *Weight* parameter indicates the weight of the patient in kilograms. *Smoke addiction* parameter defines whether the patient is a smoker or not. Smoke addition classified into four subgroups. "0" means the patient is not a smoker. "1" means the patient smokes less than five cigars per day. "2" means the patient smokes between 6 and 10 cigars per day. And "3" means the patient smokes more than 11 cigars per day. *Alcohol addiction* parameter indicates whether the patient is addicted to any kind of

Table 1 Classes for the output variable

Class name	Description	Number of patients
0.00	Patients who are not suffering from tuberculosis disease	100
0.25	Patients who have 25 % possibility of suffering from tuberculosis disease	101
0.50	Patients who have 50 % possibility of suffering from tuberculosis disease	100
0.75	Patients who have 75 % possibility of suffering from tuberculosis disease	100
1.00	Patients who are suffering from tuberculosis disease	102

Table 2 Full list of variables

Variable	Variable	Variable
1. Gender	11. Loss of appetite	21. Erythrocyte
2. Age group	12. Loss in weight	22. Haematocrit
3. Weight	13. Sweating at nights	23. Haemoglobin
4. Smoke addiction	14. Chest pain	24. Leucocyte
5. Alcohol addiction	15. Back pain	25. Number of leucocyte types
6. BCG vaccine	16. Coughing	26. Active specific lung lesion
7. Malaise	17. Haemoptysis	27. Calcific tissue
8. Arthralgia	18. Fever	28. Cavity
9. Exhaustion	19. Sedimentation	29. Pneumonic infiltration
10. Unwillingness for work	20. PPD	30. Pleural effusion
Output (0; 0.25; 0.50; 0.75; 1)		

alcohol or not. *BCG vaccine* shows the whether the patient has BCG vaccine or not.

Malaise, arthralgia, exhaustion, unwillingness for work, loss of appetite, loss in weight, sweating at nights, chest pain, back pain and coughing are binary valued parameters. They indicate whether these parameters are positive for the patient or not. Haemoptysis means coughing up blood from the respiratory tract. This parameter identifies whether the patient has haemoptysis or not. Fever is classified into three categories: “0” means normal fever value which is nearly 36.5 °C, “1” means fever value is in high ranges, and “2” means subfebrile fever value which does not exceed 38.5 °C.

In medical laboratory findings, we categorized some of blood and skin tests’ parameters. *PPD* parameter identifies whether the patient has the result of the PPD test positive (labelled as “1”) or negative (labelled as “0”). Erythrocyte is the red blood cells. They are responsible for delivering oxygen to the body tissue. We grouped this parameter into three categories. “0” means erythrocyte level is in the normal range (about 4.5–5 million per microlitre), “1” means low (which is less than 4.5 million per microlitre), and “2” means the patient has high erythrocyte level (more than 5 million per microlitre). Haematocrit is the ratio of the volume occupied by packed red blood cells to the volume of the whole blood. We also grouped this parameter into three categories. “0” means the patient has normal haematocrit percentage (about 40–45 %), “1” means low (less than 40 %), and “2” means the patient has high haematocrit percentage (more than 45 %). Haemoglobin is the iron-containing oxygen-transport metalloproteinase in the red blood cells. In our parameter values, “0” means the patient has normal haemoglobin values (about 14–16 g/dl), “1” means low (less than 14 g/dl), and “2” means the haemoglobin value is considered as high (more than 16 g/dl). Leucocytes are white blood cells. They are responsible for defending the body against infections, diseases, etc. In our parameter values, “0” means the patient has normal leucocyte values (about 5,000–10,000 in 1 mm^3 blood), “1” means leucocyte values are low (less than 5,000 in 1 mm^3 blood), and “2” means the leucocyte value is considered as high (10,000 in 1 mm^3 blood). Number of leucocyte type parameter shows the density of leucocytes, if there is a normal density (labelled as “0”) or a lymphocytic density (labelled as “1”) or macrophage density (labelled as “2”). Sedimentation parameter is a measure of the settling of red blood cells in a tube of blood during 1 h. It is grouped into three categories: “0” means normal sedimentation level (0–15 mm/h), “1” means moderately high sedimentation level (16–40 mm/h), and “2” means extremely high sedimentation level (more than 41 mm/h).

In radiological findings, active specific lung lesion parameter indicates whether there is a radiological proof of a tuberculosis lung lesion on patients’ lung. Calcific tissue shows that whether the patient has had tuberculosis disease before. If this parameter is positive, it indicates that the patient has had tuberculosis disease at least once. Cavity parameter states whether there are opening-like lesions on the patient’s lung or not. Positive value means those kinds of lesions exist on lung. Pneumonic infiltration parameter is positive if a pneumonia-like lesion is seen on the chest X-ray of patient. Pleural effusion means the accumulation of excessive pleural fluid in the pleura. This parameter is positive if such a thing is seen on the chest X-ray of patient. Table 3 lists the input variables with their data types and data domains.

2.1.2 *Feature selection for data mining methods*

Attribute ranking function is applied using information gain ranking filter in WEKA platform before applying ANFIS and rough set methods. WEKA and ROSETTA are both freeware data mining tools which are distributed freely. Before generating ANFIS model, attribute ranking function is applied using information gain ranking filter in WEKA [19] platform. InfoGainAttributeEval algorithm

Table 3 List of types and domain values of variables

Variable name	Data type	Acceptable values
Gender	Boolean	Female = 0, male = 1
Age group	Integer	18–24 = 1, 25–32 = 2, 33–40 = 3, 41–45 = 4, 46–51 = 5, 52–57 = 6, 58+ = 7
Weight	Integer	40+
Smoking addiction	Integer	None = 0, little (<5 items) = 1, moderate (6–10 items) = 2, very much (11+ items) = 3
Alcohol addiction	Boolean	No = 0, yes = 1
BCG vaccine	Boolean	No = 0, yes = 1
Malaise	Boolean	No = 0, yes = 1
Arthralgia	Boolean	No = 0, yes = 1
Exhaustion	Boolean	No = 0, yes = 1
Unwillingness for work	Boolean	No = 0, yes = 1
Loss of appetite	Boolean	No = 0, yes = 1
Loss in weight	Boolean	No = 0, yes = 1
Sweating at nights	Boolean	No = 0, yes = 1
Chest pain	Boolean	No = 0, yes = 1
Back pain	Boolean	No = 0, yes = 1
Coughing	Integer	No = 0, yes = 1, with mucous = 2
Haemoptysis	Boolean	No = 0, yes = 1
Fever	Integer	Normal = 0, high = 1, subfebrile = 2
Sedimentation	Integer	Normal = 0, moderate = 1, high = 2
PPD	Boolean	Negative = 0, positive = 1
Erythrocyte	Integer	Normal = 0, low = 1, high = 2
Haematocrit	Integer	Normal = 0, low = 1, high = 2
Haemoglobin	Integer	Normal = 0, low = 1, high = 2
Leucocyte	Integer	Normal = 0, low = 1, high = 2
Number of leucocyte types	Integer	Normal = 0, lymphocytic dense = 1, macrophage dense = 2
Active specific lung lesion	Boolean	No = 0, yes = 1
Calcific tissue	Boolean	No = 0, yes = 1
Cavity	Boolean	No = 0, yes = 1
Pneumonic infiltration	Boolean	No = 0, yes = 1
Pleural effusion	Boolean	No = 0, yes = 1
Output (bacteria existence possibility)	Float	0 % = 0, .25 % = 0.25, .5 % = 0.5, .75 % = 0.75, 100 % = 1

that we used in WEKA evaluates attributes by measuring their information gain with respect to the class. In the process of algorithm, numeric attributes are firstly discretized by using MDL-based discretization method [19]. By applying this function, we chose the most critical parameters that will affect the fuzzy model mostly. Table 4 shows the ranking result for each variable: The variables which are ranked less than 10 % were eliminated. According to this reduction in the data set, BCG vaccine, arthralgia, chest pain, smoking addiction, gender, malaise, coughing, back pain, alcohol addiction and pleural effusion variables were ignored.

The distribution of patients by age groups is plotted in Fig. 1. According to the plot, we can see that the vast majority of the patients are older than 58 years, whereas the patients who are aged between 18 and 24 years are the smallest group. Figure 1 shows the distribution of patient count over each of the age groups.

2.2 Diagnosis of tuberculosis by using adaptive neuro fuzzy inference system (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

Table 4 Importance of variables to diagnosis tuberculosis

Rank percentage	Variable	Rank percentage	Variable
0.70740	Active specific lung lesion	0.11917	Exhaustion
0.55116	Calcific tissue	0.11589	Unwillingness for work
0.48265	Number of leucocyte types	0.11027	Haemoglobin
0.43528	Weight	0.11027	Haematocrit
0.38664	Fever	0.10775	Erythrocyte
0.37107	Age group	0.09271	BCG vaccine
0.35686	PPD	0.04021	Arthralgia
0.31945	Sweating at nights	0.03608	Chest pain
0.31389	Leucocyte	0.03187	Smoking addiction
0.21179	Loss in weight	0.03029	Gender
0.21131	Haemoptysis	0.02764	Malaise
0.18745	Cavity	0.02534	Coughing
0.17851	Sedimentation	0.01626	Back pain
0.15977	Loss of appetite	0.01276	Alcohol addiction
0.13992	Pneumonic infiltration	0.00459	Pleural effusion

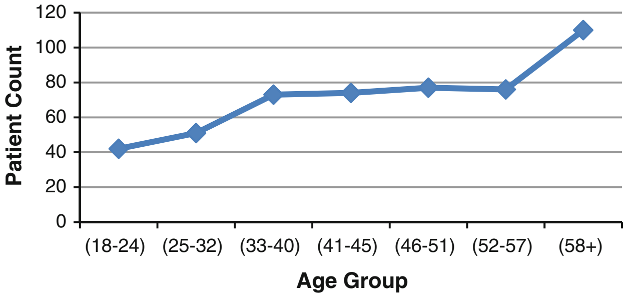

Fig. 1 Distribution of patients by their age groups

that process, fuzzy rules and fuzzy inference engine are applied to fuzzy domain [17, 20]. The obtained result is then transformed back to arithmetical domain by using defuzzifiers. Gaussian 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 linear 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 [21, 22]. Figure 2 shows a first-order Sugeno fuzzy model. The following is a typical fuzzy rule for a Sugeno-type fuzzy system:

$$\text{If } x \text{ is A and } y \text{ is B then } x = f(x, y)$$

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

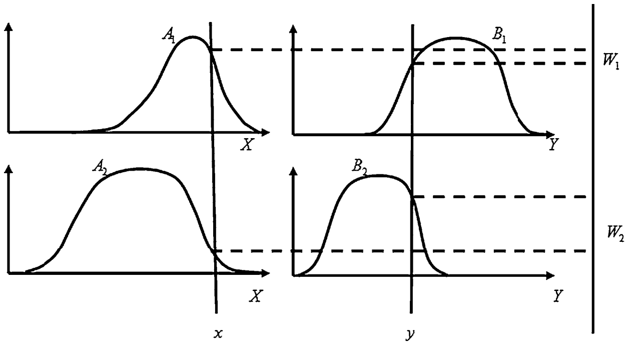

Fig. 2 First-order Sugeno fuzzy model

that are having $f(x, y)$ as a first-order polynomial. This model was originally proposed in [21, 22]. We use zero-order Sugeno fuzzy model for cases where f is constant. This can be called as a special case for Mamdani fuzzy inference system [23]. In this case, a fuzzy singleton is defined for each rule's resultant. Or, this can be also called as a special case for Tsukamoto's fuzzy model [24]. In this case, a membership function of a step function is defined where it is centred at the constant for each rules' consequent. Additionally, a radial basis function network under certain minor constraints is functionally correlative to a zero-order Sugeno fuzzy model [17]. Let us scrutinize a first-order Sugeno fuzzy inference system having two rules:

$$\text{Rule 1 : If } X \text{ is } \mathrm{A}_1 \text{ and } Y \text{ is } \mathrm{B}_1, \text{ then} f_1 = p_1x + q_1y + r_1$$

$$\text{Rule 2 : If } X \text{ is } \mathrm{A}_2 \text{ and } Y \text{ is } \mathrm{B}_2, \text{ then} f_2 = p_2x + q_2y + r_1$$

In the following figure, the fuzzy reasoning system is illustrated in a shortened form [25]. In order to bypass excessive computational complexity in the process of defuzzification, only weighted averages are used.

$$\begin{array}{l} f_1 = p_1x + q_1y + r_1 \\ f_2 = p_2x + q_2y + r_2 \end{array} \Rightarrow \begin{array}{l} t = \frac{w_1+f_1+w_2f_2}{w_1+w_2} \\ \quad = \overline{w_1}f_1 + \overline{w_2}f_2 \end{array} \tag{1}$$

On the previous figure (Fig. 3), we see a fuzzy reasoning system. This system generates an output which is shown as f. To generate this output, system accepts an input vector [x, y]. The output is calculated by computing each rule's weighted average. Those weights are achieved from the product of the membership grades in the assumption part. Using adaptive networks, which are bound with the fuzzy model, can compute gradient vectors. This computation is very helpful for learning the Sugeno fuzzy model. The resultant network is called as adaptive neuro fuzzy inference system.

The learning algorithm that ANFIS uses contains both gradient descent and the least-squares estimate. This algorithm runs over and over till an acceptable error is reached. Running process of each iteration has two phases: forward step and backward step. In forward step, linear least-squares estimate method is used for obtaining consequent parameters and precedent parameters are corrected. In backward step, fixing of consequent parameters is done. Gradient descent method is used for updating precedent parameters. And also, the output error is back-propagated through network.

It is extremely important that the number of training epochs, the number of membership functions and the number of fuzzy rules hold a critical position in the designing of ANFIS. Adjusting of those parameters is highly crucial for the system because it may lead system to over-fit the data or will not be able to fit the data. This adjusting is made by a hybrid algorithm combining the least-squares method and the gradient descent method with a mean square error method. The lesser difference between ANFIS output and the actual objective means a better (more accurate) ANFIS system. So we tend to reduce the training error in the training process.

2.2.1 *Layers of ANFIS algorithm*

A brief summary of six layers of the ANFIS algorithm is shown below. Each layer is described, and necessary formulas are stated.

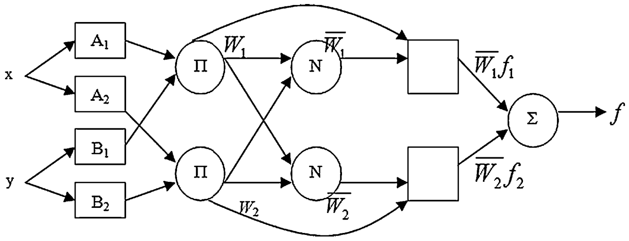

Fig. 3 ANFIS Architecture

Layer 0: It consists of plain input variable set.

Layer 1: Each node in this layer generates a membership grade of a linguistic label. For instance, the node function of the ith node may be a generalized bell membership function:

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \tag{2}$$

where x is the input to node i; A_i is the linguistic label (small, large, etc.) associated with this node; and $\{a_i, b_i, c_i\}$ is the parameter set that changes the shapes of the membership function. Parameters in this layer are referred to as the premise parameters.

Layer 2: The function is a T-norm operator that performs the firing strength of the rule, e.g., fuzzy conjunctives AND and OR. The simplest implementation just calculates the product of all incoming signals.

$$w_i = \mu A_i(x)\mu B_i(y), \quad i = 1, 2. \tag{3}$$

Layer 3: Every node in this layer is fixed and determines a normalized firing strength. It calculates the ratio of the jth rule's firing strength to the sum of all rules firing strength.

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \tag{4}$$

Layer 4: The nodes in this layer are adaptive and are connected with the input nodes (of layer 0) and the preceding node of layer 3. The result is the weighted output of the rule j.

$$\bar{w}_i f_i = \bar{w}_i(p_ix + q_iy + r_i) \tag{5}$$

where $\bar{w}_i$ is the output of layer 3, and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer are referred to as the consequent parameters.

Layer 5: This layer consists of one single node which computes the overall output as the summation of all incoming signals.

$$\text{Overall Output} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6}$$

The constructed adaptive network in Fig. 2 is functionally equivalent to a fuzzy inference system in Fig. 1. The basic learning rule of ANFIS is the back-propagation gradient descent [26], which calculates error signals (the derivative of the squared error with respect to each node's output) recursively from the output layer backward to the input nodes. This learning rule is exactly the same as the back-propagation learning rule used in the common feed-forward neural networks [17, 20, 25, 27].

2.3 Diagnosis of tuberculosis by using rough set algorithm

Rough set theory is firstly brought up by Zdzisław I. Pawlak who is a mathematician. This methodology is used for finding which attributes separate one class or classifying from another [18]. To do this, rules must be generated on a training data set. Then, those generated rules should be applied to a test data set by using rough set classification methods in order to accomplish the necessary classification task.

Approximation is a crucial concept in rough sets theory. Below, we see the approximations and their types in rough sets theory [28]:

(a) Lower Approximation (B″): Lower approximation defines the domain objects if those objects are known in which subset of interest they belong to. The Lower Approximation Set of a set X with regard to R is the set of all of objects, which certainly can be classified with X regarding R, that is, set B″.

(b) Upper Approximation (B*): This approximation defines the objects which are possibly classified among to the subset of interest. The Upper Approximation Set of a set X regarding R is the set of all of the objects which can be possibly classified with X regarding R, that is, set B*.

(c) Boundary Region (BR): If objects that of a set X regarding R is the set of all the objects, which cannot be classified neither as X nor −X regarding R, then this is called as boundary region. If the boundary region set is empty, then this kind of set is called as a "Crisp" set. If boundary region set is not empty, then it is call as a "Rough" set. Boundary region is calculated by $BR = B^{*} - B''$.

Let a set $X \subseteq U$, B be an equivalence relation and a knowledge base $K = (U, B)$. According to the definitions above, two subsets can be associated as follows:

1. B - lower: $B'' = \cup\{Y \in U/B : Y \subseteq X\}$
2. B - upper: $B^{*} = \cup\{Y \in U/B : Y \cap X \neq \emptyset\}$
   Likewise, POS(B), BN(B) and NEG(B) can be defined as follows [29].

3. $POS(B) = B'' \Rightarrow$ certainly member of X
4. $NEG(B) = U - B^{*}$
   $\Rightarrow$ certainly non - member of X
5. $BR(B) = B^{*} - B'' \Rightarrow$ possibly member of X

(d) Quality Approximation: This approximation is acquired by lower and upper approximation. The coefficient which is used in measuring the quality value is represented by αB(X). In this representation, X stands for a set of objects or registrations regarding B. In the following, the two coefficients which are used in quality approximation are declared.

- Imprecision coefficient is calculated by:

$$\alpha B(X) = |B''(X)| / |B*(X)| \tag{7}$$

- Upper and lower quality coefficients are calculated by:

$$\alpha B(B*(X)) = |B*(X)| / |A| \,(\text{upper approximation}) \tag{8}$$

$$\alpha B(B''(X)) = |B''(X)| / |A| \,(\text{lower approximation}) \tag{9}$$

Based on rough set theory, creating Rough Neural Networks is introduced by Lingras [30]. According to Lingras, every neuron in a Rough Neural Network consists of two pairs. One pair is called as upper neuron (for upper bound value) and the other is called as lower neuron (for lower bound value). The information can be changed between those two neurons as well as other rough neurons.

The rough neurons can use rough patterns. Rough patterns are built on rough values. Basically, rough values contain an upper and a lower value. In fact, this is a definition for a value range. So this type of values can be used to represent variables such as weight, age, etc. [31].

2.4 Implementing ANFIS and rough set algorithms

Rough set and ANFIS methods were used for predicting the existence of mycobacterium tuberculosis. As mentioned in the data set preparation phase, 503 different patient records were used for creating and testing ANFIS and rough set models. MATLAB's Fuzzy Logic Toolbox is used for ANFIS method. For implementing the rough set algorithm, ROSETTA [32] software was used.

3 Results and discussion

In this section, the results which were obtained from the generated data mining models are explained in detail. The classification sensitivity, specificity, precision, correctness and R.M.S.E values are given in Table 5.

As it is listed in Table 5, according to the test results of the methods, ANFIS generated sensitivity, specificity, precision and correctness values as 0.95, 0.97, 0.88 and 0.97, respectively. Rough set method produced sensitivity, specificity, precision and correctness values as 0.93, 0.92, 0.77 and 0.92, respectively. If we look to the R.M.S.E values of the methods, ANFIS has a score of 0.18 and Rough set has a score of 0.22. The plot of testing data and FIS output is displayed on Fig. 4.

Table 5 Benchmarking the accuracy of the methods

Method name	Sensitivity	Specificity	Correctness	R.M.S.E
ANFIS	0.95	0.97	0.97	0.18
Rough set	0.93	0.92	0.92	0.22

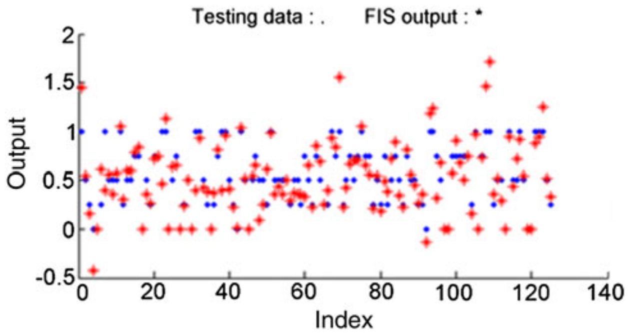

Fig. 4 ANFIS testing error plot

Two similar studies were conducted by Er et al. [8, 33]. In [8], authors constructed a multi-layer neural network model with a classification accuracy of 0.90 for tuberculosis disease, and in [33], authors constructed an artificial immune system with a classification accuracy of 0.90 for tuberculosis disease.

ANFIS generated 8 rules which explain the relationship between input parameters and output:

Rule 1: *[4 54 1 1 0 1 1 1 2 2 1 1 1 1 0 1 1 1 0 0] [1]*

If Age Group = 4 and Weight = 54 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 1 and Sweating at Nights = 1 and Haemoptysis = 1 and Fever = 2 and Sedimentation = 2 and PPD = 1 and Erythrocyte = 1 and Haematocrit = 1 and Haemoglobin = 1 and Leucocyte = 0 and Number of Leucocyte Types = 1 and Active Specific Lung Lesion = 1 and Calcific Tissue = 1 and Cavity = 0 and Pneumonic Infiltration = 0 then Output is 1.

For the inputs for rule 1, system states that this patient belongs to cluster 1 which means the patient is suffering from tuberculosis disease. If we take a closer look to the parameters, we see that active specific lung lesion and existence of calcific tissue parameters are both positive. Calcific tissue existence is a proof that patient has had tuberculosis disease at least once before. Sedimentation value is high; subfebrile fever and PPD test result is also positive. Number of leucocyte types parameters shows that there is a lymphocytic density, and patient is suffering from haemoptysis. Exhaustion, unwillingness for work, sweating at nights and loss in weight parameters also support this output.

Rule 2: *[4 54 1 1 0 1 1 1 2 2 1 1 1 1 0 0 1 1 0 0] [1]*

If Age Group = 4 and Weight = 54 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 1 and Sweating at Nights = 1 and Haemoptysis = 1 and Fever = 2 and Sedimentation = 2 and PPD = 1 and Erythrocyte = 1 and Haematocrit = 1 and Haemoglobin = 1 and Leucocyte = 0 and Number of Leucocyte Types = 0 and Active Specific Lung Lesion = 1 and Calcific Tissue = 1 and Cavity = 0 and Pneumonic Infiltration = 0 then Output is 1.

Rule 2 is similar to rule 1. The strongest parameters are positive. The supportive parameters such as exhaustion, unwillingness for work, sweating at nights and loss in weight are positive too. The only change in this case is number of leucocyte type's parameters. These parameters have a value within normal range. But the other parameters still affect the output as 1 which means the patient is suffering from tuberculosis disease.

Rule 3: *[4 54 1 1 0 1 1 1 2 2 1 1 1 1 0 1 1 0 0 0] [1]*

If Age Group = 4 and Weight = 54 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 1 and Sweating at Nights = 1 and Haemoptysis = 1 and Fever = 2 and Sedimentation = 2 and PPD = 1 and Erythrocyte = 1 and Haematocrit = 1 and Haemoglobin = 1 and Leucocyte = 0 and Number of Leucocyte Types = 1 and Active Specific Lung Lesion = 1 and Calcific Tissue = 0 and Cavity = 0 and Pneumonic Infiltration = 0 then Output is 1.

In rule 3, the parameters are mostly similar to rule 1 and rule 2. The only difference is existence of calcific tissue is negative in this rule. This means that this patient is not suffered from tuberculosis disease before. But the rest of the parameters support that patient is suffering from tuberculosis now. So the output is stated as cluster 1 by system.

Rule 4: *[4 54 1 1 0 1 1 1 0 2 1 1 1 1 0 1 1 1 0 0] [1]*

If Age Group = 4 and Weight = 54 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 1 and Sweating at Nights = 1 and Haemoptysis = 1 and Fever = 0 and Sedimentation = 2 and PPD = 1 and Erythrocyte = 1 and Haematocrit = 1 and Haemoglobin = 1 and Leucocyte = 0 and Number of Leucocyte Types = 1 and Active Specific Lung Lesion = 1 and Calcific Tissue = 1 and Cavity = 0 and Pneumonic Infiltration = 0 then Output is 1.

Rule 4 covers cases where patient has a fever value within normal ranges. This is a normal possibility in real life too. When we look into other parameters, we see that active specific lung lesion and existence of calcific tissue parameters are both positive. PPD is positive and patient has a high level of sedimentation. Patient is also having haemoptysis. And the supportive parameters such as exhaustion, unwillingness for work, sweating at nights and loss in weight are positive. So system classifies this kind of patient to cluster 1 which means there is an active tuberculosis disease.

Rule 5: *[2 69 1 1 0 0 0 0 1 1 0 0 0 0 2 2 0 0 0 1] [0]*

If Age Group = 2 and Weight = 69 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 0 and Sweating at Nights = 0 and Haemoptysis = 0 and Fever = 1 and Sedimentation = 1 and PPD = 0 and Erythrocyte = 0 and Haematocrit = 0 and Haemoglobin = 0 and Leucocyte = 2 and Number of Leucocyte Types = 2 and Active Specific Lung Lesion = 0 and Calcific Tissue = 0 and Cavity = 0 and Pneumonic Infiltration = 1 then Output is 0.

Rule 5 has an output of 0 which means that patient is not suffering from tuberculosis disease. Exhaustion and unwillingness for work is positive. Fever is in high values. Sedimentation is moderately high, leucocyte value is high and macrophage density is spotted in number of leucocyte types. Pneumonic infiltration is also positive. These parameters indicate that patient is not having tuberculosis disease. He/she is most probably having another disease such as pneumonia.

Rule 6: *[1 71 1 1 0 0 0 0 1 2 0 0 0 0 2 2 0 0 0 0] [0]*

If Age Group = 1 and Weight = 71 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 0 and Sweating at Nights = 0 and Haemoptysis = 0 and Fever = 1 and Sedimentation = 2 and PPD = 0 and Erythrocyte = 0 and Haematocrit = 0 and Haemoglobin = 0 and Leucocyte = 2 and Number of Leucocyte Types = 2 and Active Specific Lung Lesion = 0 and Calcific Tissue = 0 and Cavity = 0 and Pneumonic Infiltration = 0 then Output is 0.

In rule 6, exhaustion and unwillingness for work is positive. Fever is in high values. Sedimentation value is also high. There is no active specific lung lesion and calcific tissue existence is also negative. Leucocyte parameter is high and macrophage density is spotted in number of leucocyte types. But pneumonic infiltration is not positive. This indicates that patient is not having tuberculosis disease but he/she is most probably suffering from acute bronchitis.

Rule 7: *[2 68 1 1 1 0 0 0 1 2 0 0 0 0 2 2 0 0 0 1] [0]*

If Age Group = 2 and Weight = 68 and Exhaustion = 1 and Unwillingness for Work = 1 and Loss of Appetite = 1 and Loss in Weight = 0 and Sweating at Nights = 0 and Haemoptysis = 0 and Fever = 1 and Sedimentation = 2 and PPD = 0 and Erythrocyte = 0 and Haematocrit = 0 and Haemoglobin = 0 and Leucocyte = 2 and Number of Leucocyte Types = 2 and Active Specific Lung Lesion = 0 and Calcific Tissue = 0 and Cavity = 0 and Pneumonic Infiltration = 1 then Output is 0.

Parameters of rule 7 are similar to rule 6 for most of the values. In this case, the only difference is pneumonic infiltration value which is positive. This patient is in cluster 0 and he/she is probably having another disease such as pneumonia.

Rule 8: *[3 55 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] [0]*

If Age Group = 3 and Weight = 55 and Exhaustion = 0 and Unwillingness for Work = 0 and Loss of Appetite = 0 and Loss in Weight = 0 and Sweating at Nights = 0 and Haemoptysis = 0 and Fever = 0 and Sedimentation = 0 and PPD = 0 and Erythrocyte = 0 and Haematocrit = 0 and Haemoglobin = 0 and Leucocyte = 0 and Number of Leucocyte Types = 0 and Active Specific Lung Lesion = 0 and Calcific Tissue = 0 and Cavity = 0 and Pneumonic Infiltration = 0 then Output is 0.

If we look to the input parameters of the last rule, Rule 8, it is obvious that this case is not suffering from tuberculosis disease. All of the major parameters that indicate tuberculosis are having normal values.

The rules above indicate two distinct classes for the trained ANFIS model. These classes are 0 and 1. An instance in class 0 means that patient is not suffering from tuberculosis disease and class 1 means that patient is suffering from tuberculosis disease. Each rule is represented by a vector which contains input values for the model. As stated in the data preparation section, each input variable has a different ranking on affecting the output. For instance, active specific lung lesion parameter has a ranking of 70 %, calcific tissue existence parameter has a ranking of 55 %, number of leucocyte types parameter has a ranking of 48 % and weight of the patient parameter has a ranking value of 43 %. Those four parameters are the strongest ones among other parameters. Especially active specific lung lesion parameter shows great importance. Two of the strongest input parameters which are active specific lung lesion and calcific tissue existence are plotted versus output in the surface diagram that is displayed in Fig. 5.

Figure 6 shows the surface plot of patient weight and age group parameters versus output. The surfaces in Figs. 5 and 6 show us the change of output according to the two given input parameter pairs. Figure 7 shows the plot of age group versus output. The curve in Fig. 7 directly shows the change of output according to the given age group parameter.

Rough set model generated 30 rules. The following is the full list of rules for the generated Rough set model:

Rule 1: Leucocyte(2) ⇒ Output(0.00) OR Output(1.00)

Rule 2: Weight([76, 95]) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 3: Weight([0, 45]) ⇒ Output(0.75) OR Output(1.00)
Rule 4: Fever(1) ⇒ Output(0.00)
Rule 5: Age interval(2) ⇒ Output(0.00) OR Output(0.75) OR Output(1.00)
Rule 6: Age interval(1) ⇒ Output(0.00) OR Output(1.00)
Rule 7: Age interval(6) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 8: Age interval(3) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 9: Age interval(7) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)

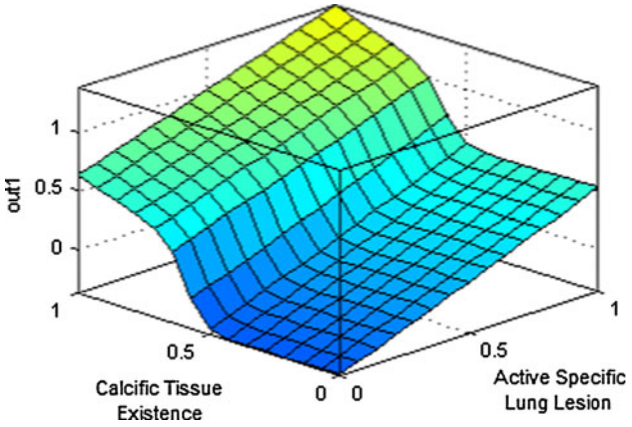

Fig. 5 Surface plot of active specific lung lesion and calcific tissue existence parameters versus output

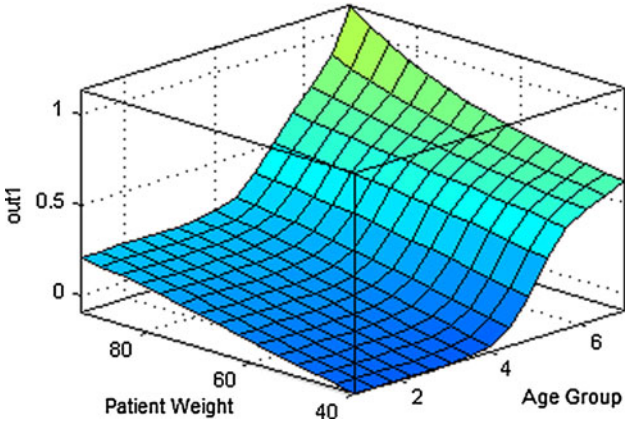

Fig. 6 Surface plot of patient weight and age group parameters versus output

Rule 10: Age interval(4) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 11: Age interval(5) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 12 Pneumonic infiltration(1) ⇒ Output(0.00)
Rule 13 Calcific Tissue(0) ⇒ Output(0.00) OR Output(1.00)
Rule 14: Unwillingness for work(1) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 15: Unwillingness for work(2) ⇒ Output(1.00)
Rule 16: Sedimentation(0) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 17: Sedimentation(2) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 18: Weight(1) AND Sweating at night(0) ⇒ Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 19: Age interval(3) AND Loss of appetite(1) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75)
Rule 20: Number of leucocyte types(1) ⇒ Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 21: Weight(1) ⇒ Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 22: Erythrocyte(1) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 23: Weight(1) AND Sedimentation(1) ⇒ Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)

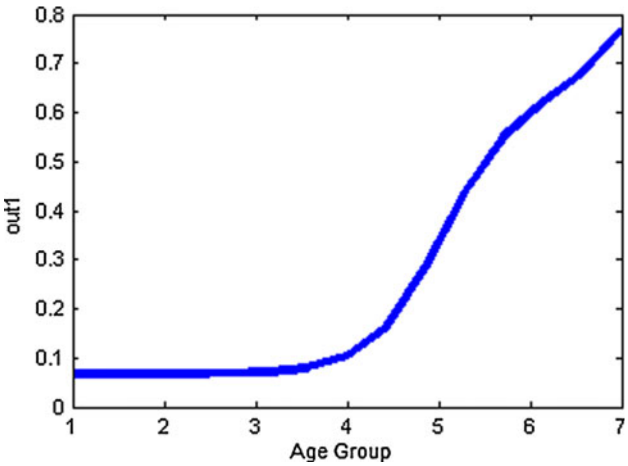

Fig. 7 Plot of age group versus output

Rule 24: Age interval(6) AND Exhaustion(0) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75)
Rule 25: Age interval(4) AND Weight(1) ⇒ Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 26: Age interval(7) AND Fever(2) ⇒ Output(0.00) OR Output(0.25) OR Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 27: Haemoptysis(1) ⇒ Output(0.50) OR Output(0.75) OR Output(1.00)
Rule 28: Haemoptysis(2) ⇒ Output(0.75)
Rule 29: Active specific lung lesion(1) ⇒ Output(1.00)
Rule 30: Cavity(1) ⇒ Output(1.00)

According to these generated rules, it is clear that rough set algorithm uses very few combinations of input parameters for classification task.

4 Conclusions

Data mining techniques are used widely in biomedical area. There are many studies performed using different techniques. Each of these techniques has pros and cons. The main aim of this study is developing a data mining solution which makes diagnosis of tuberculosis as accurate as possible and helps deciding whether it is reasonable to start tuberculosis treatment on suspected patients without waiting the exact medical test results or not.

In order to create the desired data mining model, a data set of 503 records each having 30 attributes was used. For selecting the most effective attributes in the data set, ranking algorithm (InfoGainAttributeEval with Ranker) was applied. According to the ranking result, 10 attributes were removed. The removed attributes were the ones which were ranked less than 0.10. So, they were not having much importance on the data set at all.

Each of the attributes on the data set represents a value for the patient such as gender, loss of appetite, age group of patient, loss in weight, total weight of patient, smoke addiction level, chest pain level, etc. The model that was developed distinguishes the probability class of the patient using these attributes whether he/she is currently suffering from tuberculosis disease or not.

ANFIS model classifies the patients with a correctness of 97 % where rough set method makes the same classification with a correctness of 92 %.

Benchmarking values indicate that the ANFIS model that was developed classifies the instances with a very acceptable result when comparing with rough set method. ANFIS's generated rules' integrity and consistency is checked by comparing each rule with real case inputs and outputs. If we compare the generated rules of ANFIS and rough set algorithms, we see that ANFIS generated more generalized rules for cases. The rules which are generated by rough set algorithm are much more specific and mostly focused on single input parameters. So this reduces accuracy of rough set. On the other hand, our findings indicate that rough set method is also having RMSE result within an acceptable range. But ANFIS has a better RMSE score.

According to the findings of this study, ANFIS is an accurate and reliable method when comparing with rough set method for classification of tuberculosis patients.

References

1. Davidson S (1999) Davidson's principles and practice of medicine. Churchill Livingstone, London
2. Harrison TR (1999) Harrison's principles of internal medicine. McGraw-Hill Education, New York
3. Özlü T, Metintaş M, Ardıç S (2008) Akciğer Hastalıkları Temel Bilgiler. Poyraz Tıbbi Yayıncılık, Ankara
4. Bozkurt H, Türkkanı MH, Musaonbaşıoğlu S, Güllü Ü, Baykal F, Hasanoğlu C, Özkara Ş (2009) Türkiye'de Verem Savaşı 2009 Raporu. T.C. Sağlık Bakanlığı, Ankara
5. Pena A, Domínguez R, Medel J (2009) Educational data mining: a sample of review and study case. World J Educ Technol 1(2):118–139
6. Bakar AA, Febriyani F (2007) Rough neural network model for tuberculosis patient categorization. In: Proceedings of the international conference on electrical engineering and informatics, Indonesia, pp 765–768
7. Sánchez MA, Uremovich S, Acrogliano P (2009) Mining tuberculosis data. In: Berka P, Rauch J, Zighed DA (eds) Data mining and medical knowledge management: cases and applications. Medical Information Science Reference, New York, pp 332–349
8. Er O, Yumusak N, Temurtas F (2010) Chest diseases diagnosis using artificial neural networks. Expert Syst Appl 37(12):7648–7655
9. Park M, Kang B, Jin SJ, Luo S (2009) Computer aided diagnosis system of medical images using incremental learning method. Expert Syst Appl 36(3):7242–7251
10. Yang HH, Wu CL (2009) Rough sets to help medical diagnosis—evidence from a Taiwan's clinic. Expert Syst Appl 36(5):9293–9298
11. Yan H, Jiang Y, Zheng J, Peng C, Li Q (2006) A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Syst Appl 30(2):272–281
12. Karahoca A, Karahoca D, Kara A (2009) Diagnosis of diabetes by using adaptive neuro fuzzy inference systems. In: Soft computing, computing with words and perceptions in system analysis, decision and control, Famagusta, pp 1–4
13. Monzon JE, Pisarello MI (2005) Cardiac beat classification using a fuzzy inference system. In: Proceedings of the 2005 IEEE engineering in medicine and biology 27th annual conference, Shanghai, pp 5582–5584
14. Gómez C, Hornero R, Abásolo D, Fernández A, Escudero J (2009) Analysis of MEG background activity in alzheimer's disease using nonlinear methods and ANFIS. Ann Biomed Eng 37(3):586–594
15. Shlomi T, Cabili MN, Ruppin E (2009) Predicting metabolic biomarkers of human inborn errors of metabolism. EMBO and Macmillan Publishers Limited, Israel Institute of Technology, Haifa

16. Gören S, Karahoca A, Onat FY, Gören Z (2008) Prediction of cyclosporine A blood levels: an application of the adaptive-network-based fuzzy inference system (ANFIS) in assisting drug therapy. Eur J Clin Pharmacol 64(8):807–814
17. Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybernet 23(3):665–685
18. Pawlak Z (1984) Rough classification. Int J Man-Machine Stud 20(5):469–483
19. Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques. Morgan Kaufmann Publishers, San Fransisco
20. Jang JS (1992) Self-learning fuzzy controllers based on temporal back propagation. IEEE Trans Neural Netw 3(5):714–723
21. Sugeno M, Kang GT (1988) Sturcture identification of fuzzy model. Fuzzy Sets Syst 28(1):15–33
22. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its application to modeling and control. IEEE Trans Syst Man Cybern 15(1):116–132
23. Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man-Machine Stud 7(1):1–13
24. Tsukamato Y (1979) An approach to fuzzy reasoning method. In: Gupta MM, Ragade RK, Yager RR (eds) Advances in fuzzy set theory and applications. Elsevier Science Ltd, Amsterdam, pp 137–149
25. Jang JS (1996) Input selection for ANFIS learning. In: Proceedings of the IEEE international conference on fuzzy systems, New Orleans, pp 1493–1499
26. Werbos P (1974) Beyond regression, new tools for prediction and analysis in the behavioural sciences. PhD Thesis, Harvard University
27. Chiu SL (1997) Extracting fuzzy rules from data for function approximation and pattern classification. In: Dubois D, Prade H, Yager R (eds) Fuzzy information engineering: a guided tour of applications. Wiley, New York
28. Ponce J, Karahoca A (2009) Data mining and knowledge discovery in real life applications. IN-TECH
29. Pawlak Z (1991) Rough sets: theoretical aspects of reasoning about data. Kluwer, Norwell
30. Lingras P (1996) Rough neural networks. In: Proceeding of the 6th international conference on information processing and management of uncertainty in knowledge-based systems, Granada, pp 1445–1450
31. Lingras P (1998) Comparison of neofuzzy and rough neural networks. Inf Sci 110(3–4):207–215
32. Øhrn A (1999) Discernibility and rough sets in medicine: tools and applications. PhD Thesis. Norwegian University of Science and Technology, Trondheim. ISBN 82-7984-014-1
33. Er O, Yumusak N, Temurtas F (2012) Diagnosis of chest diseases using artificial immune system. Expert Syst Appl 39(2):1862–1868