Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 5 (2017) pp. 033-039


# Tuberculosis Diagnosis Using Adaptive Neuro-Fuzzy Inference Systems

**Semagn Tiruneh Sisay[1], Niketa Gandhi[2]**

[1] IT Doctoral Program, Addis Ababa University
Addis Ababa, Ethiopia
*tsemagn@gmail.com*

[2] Machine Intelligence Research Labs (MIR Labs)
Scientific Network for Innovation and Research Excellence,
P.O. Box 2259, Auburn, Washington 98071, USA
*niketa@gmail.com*

***Abstract:*** Tuberculosis (TB) is one of the leading infectious diseases all over the world. TB affects millions of people every year and more than 10% of them die due to this disease. Despite the belief that it is almost under control and the availability of age-old cure effective available, TB continues to infect humankind and it remains a global emergency. The traditional methods of TB diagnosis are inaccurate and time-taking, expensive, low efficacy rates, may give false results, cannot differentiate between latent TB and active TB, and unable to differentiate drug resistant TB stages, and cannot be detect TB in case of HIV and TB co-infection due to low levels of TB bacteria. Besides, TB diagnosis in developing countries faced challenges like poor diagnosis tools, low level laboratory systems and medical facilities, and lack of data processing culture. Therefore, it is inevitable to search for new TB diagnosis techniques that give accurate results with a greater speed. This study proposes a technique for TB diagnosis using Adaptive Neuro-Fuzzy Inferential System to provide a tool for accurate, timely, and cost effective diagnosis of Tuberculosis.

***Keywords:*** artificial intelligence, adaptive neuro-fuzzy inference systems, diagnosis, expert systems, fuzzy logic, neural network, tuberculosis

## I. Introduction

Health is an area of concern to all population [1]; and it is a vital factor of human development [2]. Healthcare has direct relation with productivity, educational performance, life expectancy and savings that lead to social and political stability [2]. However, some countries face problems of critical healthcare issues, lack of basic healthcares programs, lack of the required workers, poor healthcare facilities, unsustainable of services, and backward data processing cultures [1],[3],[4]. Healthcare decisions are made based on crude estimates and substitutes for unavailable key indicators [3],[5]. Therefore, preventable diseases and premature deaths are still very high in these countries [6],[7].

Tuberculosis (TB) is one of the most common causes of death worldwide and it is a severe problem for the nations [8],[9]. Despite the belief that it is almost under control and the effective cure available, TB continues to plague humankind [8] and it remains a global emergency affecting millions [10]. World Health Organization report in 2015 estimates 9.6 million people developed TB of which 15.6 % (from which 22.7% among HIV-negative people and 73.3% among HIV-positive people) died while a report in 2016 estimates around 10.4 million people developed TB of which 13.7% (from which 28.6% among HIV-negative people and 71.4% among HIV-positive people) died [11],[12]. The reports show that Africa bears the highest burden and more than half of cases are in Sub-Saharan Africa. However, if it is detected at early stages then most of the TB deaths can be prevented.

TB is an infectious disease caused by *Mycobacterium tuberculosis*. Detection of bacteria is very important to prevent its growth and maintain human health [13]. However, traditionally TB diagnosis method, which is carried out using physical examination and laboratory test is inaccurate and time-taking, expensive, low efficacy rates, may give false results, cannot differentiate between latent TB and active TB, and unable to differentiate drug resistant TB stages, and cannot detect TB in case of HIV and TB co-infection due to low levels of TB bacteria [8],[9]. Besides, TB diagnosis in developing countries face challenges like poor diagnosis tools, low level laboratory systems and medical facilities, and lack of data processing culture [14]. This way TB remains undefined and a killing disease in Africa.

As discussed, drawbacks of the traditional diagnosis methods (examining sputum, X-ray, and culture) forced researchers to find and develop TB diagnosis techniques that give accurate results with a greater speed. The new methods are artificial intelligence (AI) tools like expert systems (ESs), artificial neural networks (ANNs), fuzzy logic (FL), neuro-fuzzy inference systems (NFISs) and genetic algorithms (GAs).

### *A. Background*

The growth of Information and Communication Technologies (ICTs) provides opportunities to confront such health disparities, social and economic inequalities, and improve human development. Since ICTs provide effective and efficient solutions to real life problems related to medicine [14], morbidity and mortality rates are reducing. Currently, data mining techniques are used for diagnosis and predicting diseases with accuracy and speed. For this study, diagnosing the existence of mycobacterium tuberculosis on patients is done by using adaptive neuro-fuzzy inference system (ANFIS).

### B. *Objective of the study*

The objective of this study is to diagnose TB by using ANFIS. The study is done using the following stages:

1. Data preprocessing
   - A. Data collection
   - B. Avoiding incomplete data
   - C. Selection of diagnosis parameters
   - D. Data processing
2. Study of neuro-fuzzy inference systems
4. Analysis of TB diagnosis results

## II. Literature Review

This section presents a review of literature on the concepts of AI and the descriptions of AI methods tools for building artificial systems.

Broadly AI is the automation of activities that are associated with human thinking [15] in order to handle complex tasks that require human intervention [8],[9]. Such tasks include reasoning, learning, decision making, programming, natural language understanding and translation, knowledge processing and representation [8],[9], [16],[17],[18]. The AI methods include artificial intelligence (AI) tools like expert systems (ESs), artificial neural networks (ANNs), fuzzy logic (FL), neuro-fuzzy inference systems (NFISs) and genetic algorithms (GAs) [33-38].

**Expert System (ES)**

Expert systems (ESs) imitate human thinking through knowledge representations and inferences to solve problems that require professional expertise [14],[19]. Within a bound domain of knowledge, ESs are capable of decision making on a level comparable in quality to human experts [19].

**Neural Networks (NN)**

Neural network (NNs) is a group of interconnected artificial neurons that mimic the properties of biological neurons**.** Artificial neural network (ANN) simulates functions of human brain or networking system aiming at processing big data simultaneously [20],[21]. In ANNs, signals reach the neurons at their axon terminals through synapses between the dendrites and axon terminals of the neuron. When the signals are strong enough to surpass a certain threshold, activation of the neuron takes place and results in the emission of a signal through its axon.

Neural network consists of three layers: the first (input) layer receives the input signals; the middle (hidden) layer propagates signals from first layer to the third (output) layer; and the third (output) layer produces the result of the process.

**Fuzzy Logic (FL)**

Fuzzy logic (FL) deals with uncertainty in knowledge with the aim of providing approximate reasoning in a fuzzy data [22]. FL theory provides a mathematical strength to capture uncertainties and it is all about relative importance of precision. FL has the ability to represent comprehensive linguistic knowledge and perform reasoning through fuzzy rules [23]. In fuzzy set theory, each element of the Membership Function (MF) has a unit value that characterizes the grade of membership of a set. Here, an element of a given set can belong to another set at varying degrees [22].

**Genetic Algorithms (GAs)**

GAs are search algorithms that use the techniques of human genetic evolution to generate optimal solutions, where candidate solutions to a problem will evolve over a period of time through competition and controlled variation [24],[25].

**Adaptive neuro-fuzzy inference system (ANFIS)**

ANFIS is a combination of ANN and FL principles into a single framework [24]. Its inference system corresponds to a set of fuzzy IF–THEN rules that have a higher capability in the learning process to approximate nonlinear functions and to adapt its environment [26],[27]. Therefore, ANFIS is considered to be a universal estimator with greater efficiency than either ANN or FL [28],[29].

A Fuzzy Inference System (FIS) has three main components [26]:

- Basic rules: where it consists of the selection of fuzzy logic rules;
- “If-Then” as a function of the fuzzy set membership; and
- Reasoning fuzzy inference techniques from basic rules to get the output.

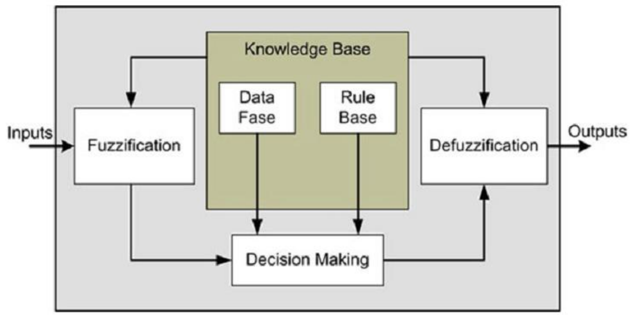


Figure 1. Fuzzy Inference System
*(Source: W. Suparta and K.M. Alhasa, 2016)*

FIS works as follows [30], [31], [32]:

- Fuzzification: To convert numerical domain into fuzzy domain;
- Applying fuzzy rules and fuzzy inference engine to fuzzy domain;
- Defuzzification: transforming the obtained result back to numerical domain;
- Summation of outputs is calculated at the last node of the system.

Adaptive network is a feed forward neural network with multiple layers. It consists of a number of adaptive nodes interconnected directly without any weight value between them. In adaptive network, each node has different functions and tasks, and the output depends on the incoming signals and parameters that are available in the node [26]. It uses supervised learning algorithm where a learning rule that was used can affect the parameters in the node and it can reduce the occurrence of errors at the output of the adaptive network [24].

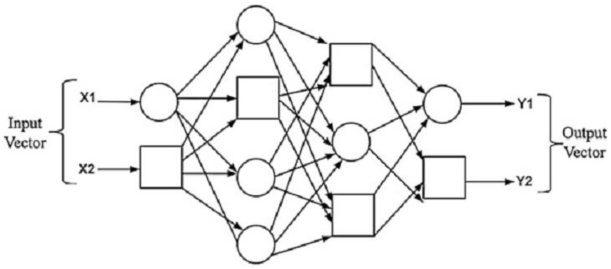


Figure 2. Fuzzy Inference System
*(Source: W. Suparta and K.M. Alhasa, 2016)*

### Architecture of ANFIS

To present the ANFIS architecture, two fuzzy if-then rules based on a first order Sugeno model are considered:

Rule 1: If *($x$ is $A_1$)* and *($y$ is $B_1$)* Then *($f_1 = p_1x + q_1y + r_1$)*
Rule 2: If *($x$ is $A_2$)* and *($y$ is $B_2$)* Then *($f_2 = p_2x + q_2y + r_2$)*

Where x and y are the inputs, $A_i$ and $B_i$ are the fuzzy sets, $f_i$ are the outputs within the fuzzy region specified by the fuzzy rule, $p_i$, $q_i$ and $r_i$ are the design parameters that are determined during the training process.

## III. Data collection and preprocessing method

In order to conduct this research, the patient's TB data taken from Arada Health Center in Addis Ababa, Ethiopia, was used. The dataset were prepared using TB diagnosis registries that register patients' name and address, examination types conducted and associated results, medications follow ups, and progress tracks. The dataset consists of TB test types with five classes and 523 samples. The class distribution is

- Class 1: Smear (Sputum test/P/Pos) = 131
- Class 2: X-ray (P/Neg) = 171
- Class 3: Culture (EP) = 200
- Class 4: Smear (Sputum test/P/Pos) and X-ray (P/Neg) = 5
- Class 5: Normal (TB Negative) = 22

Incomplete data and missed values are rejected, and only the complete data are identified and used in this study. The data are encoded into excel 2017 and later changed into a comma-separated values (CSV) file.

### a. Materials and Methods

***Preparing TB data set***

Data set was collected from Arada Health Center at Addis Ababa, Ethiopia. A total of 523 patient data was collected, which covers patients' examination reports from 01.09.2016 and 31.12.2017. As shown in Table 1, data set was divided into five different classes based on the diagnosis types. Each patient record consists of four variables directly linked to TB diagnosis, but more variables.

***Domain values of input variables***

Attributes of the data set can be categorized into four groups as follows:

- Demographics data and clinical findings: gender indicates whether the patient is male or female; age group indicates the age group that the patient belongs to. All ages are group into thirteen different classes. These classes are "10-14", "15-19", "20-24", "25-29", "30-34", "35-39", "40-44", "45-49", "50-54", "55-59", "60-64", "65-69", ">=70"
- Sputum findings: where the patient is TB positive or negative based on sputum test.
- X-ray findings: where the patient is TB positive or negative based on X-ray examination.
- Culture findings: where the patient is TB positive or negative based on culture test taken from his/her body parts.

Table 1. Output variable

| Description | Number of patients |
|---|---|
| Patients who are TB negative | 22 |
| Patients who are TB positive using sputum test only | 131 |
| Patients who are TB positive using X-ray examination only | 171 |
| Patients who are TB positive using culture test only | 200 |
| Patients who are TB positive using both sputum test and X-ray examination | 5 |

Table 2. Full list of variables

| | |
|---|---|
| Age | X-ray |
| Sex | Culture (EP) |
| Sputum | TB result |

In all types of test/examination findings, tests' parameters show that whether the patient has TB disease or not. Since the data available do not show the detail test parameters, this study categorizes the test as positive and negative generally if TB disease is present or absent respectively with respect to each test type.

***Feature selection for data mining methods***

Attribute ranking function is applied using information gain ranking filter in WEKA platform before applying ANFIS.

Table 3**.** List of types and domain variables

| Variable name | Data Type | Acceptable values |
|---|---|---|
| Gender | Boolean | Female = 0, Male = 1 |
| Age | Integer | 10-14=1, 15-19=2, 20-24=3, 25-29=4, 30-34=5, 35-39=6, 40-44=7, 45-49=8, 50-54=9, 55-59=10, 60-64=11, 65-69=12, +=70=13 |
| Sputum | Boolean | Negative = 0, Positive = 1 |

| | | |
|---|---|---|
| X-ray | Boolean | Negative = 0, Positive = 1 |
| Culture (EP) | Boolean | Negative = 0, Positive = 1 |
| TB result | Boolean | Absent = 0, Present = 1 |

***Data analysis***

This study used Weka 3.8 to process data. Weka 3.8 produces the following output:

```
Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         518               99.044 %
Incorrectly Classified Instances         5                0.956 %
Kappa statistic                          0.8669
Mean absolute error                      0.0189
Root mean squared error                  0.0973
Relative absolute error                 23.0342 %
Root relative squared error             48.4655 %
Total Number of Instances              523
```

Figure 3 (a). Data processing output

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall
                 1.000    0.227    0.990      1.000
                 0.773    0.000    1.000      0.773
Weighted Avg.    0.990    0.218    0.991      0.990
```

Figure 3 (b). Data processing output

```
F-Measure  MCC    ROC Area  PRC Area  Class
0.995      0.875  0.886     0.990     P
0.872      0.875  0.886     0.782     N
0.990      0.875  0.886     0.981
```

Figure 3 (c). Data processing output

```
=== Confusion Matrix ===

   a   b   <-- classified as
 501   0 |   a = P
   5  17 |   b = N
```

Figure 3 (d). Data processing output

***Decision table***

Table 4. Decision table for TB diagnosis using smear test, x-ray and culture examinations

| Condition | Smear Test is Positive | X-Ray is Positive | ERP is Positive | TB Result |
|---|---|---|---|---|
| Rules | No | No | No | Negative |
| | No | No | Yes | Positive |
| | No | Yes | No | Positive |
| | No | Yes | Yes | Positive |
| | Yes | No | No | Positive |
| | Yes | No | Yes | Positive |
| | Yes | Yes | No | Positive |
| | Yes | Yes | Yes | Positive |

***Decision tree***

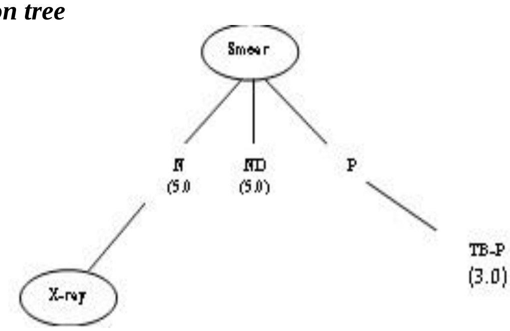


Figure 4 (a). Decision tree for TB diagnosis using smear test, x-ray and culture examinations

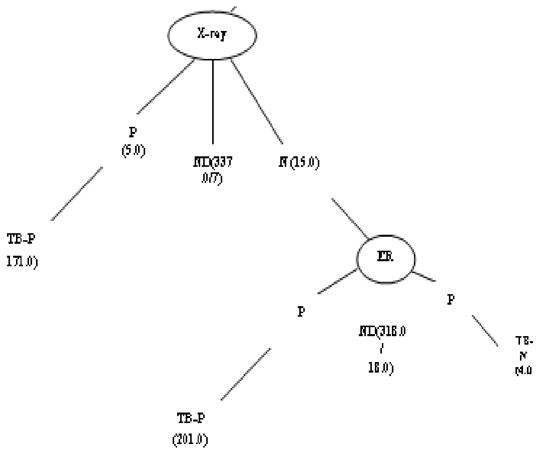


Figure 4 (b). Decision tree for TB diagnosis using smear test, x-ray and culture examinations

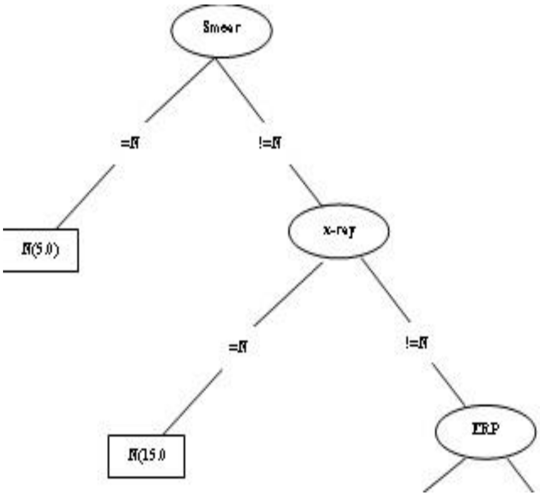


Figure 5(a). Decision tree for TB diagnosis using smear test, x-ray and culture examinations

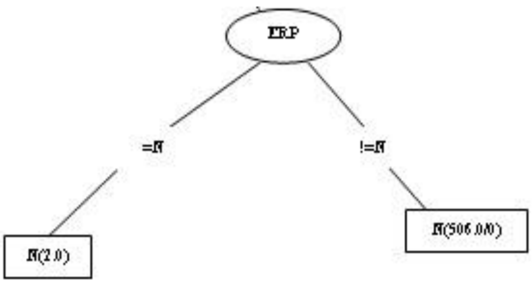


Figure 5(b). Decision tree for TB diagnosis using smear test, x-ray and culture examinations

### *Rules*

IF Smear-test is positive THEN TB-result is positive

Else IF Smear-test is negative OR x-ray is positive THEN TB-result is positive

Else IF Smear-test is negative OR ERP is positive THEN TB-result is positive

Else IF X-ray is negative OR ERP is positive THEN TB-result is positive

Else IF ERP is negative OR X-ray is positive THEN TB-result is positive

Else IF Smear-test is positive OR X-ray is positive OR ERP is positive THEN TB-result is positive

Else IF Smear-test is negative AND X-ray is negative AND ERP is negative THEN TB-result is negative.

## IV. Empirical Results

The results show that there are 518 (99.044% correctly classified instances and 5(0.956%) incorrectly classified instances. The mean absolute error is 0.0189, which is absolute difference between the forecasted value to the actual value is very low. Besides, the Kappa Statistic (0.8669) shows that there is an excellent level of agreement between the variables; with greater precision (0.990) and recall (1.00). As the confusion matrix shows, the diagnosis model is highly right with greater algorithm performance; that is, the model with no (or very little) confusion. Therefore, the model can diagnose TB with greater accuracy.

## V. Conclusions

The proposed ANFIS provides techniques of TB diagnosis that generate precise results. It aids TB diagnosis with different input parameters, and the test results are generated based on the inputs applied. This method greatly helps physicians to make decisions while conducting TB diagnosis. Besides, it can be extended for any number of inputs.

Since ANFIS makes accurate and fast results on TB diagnosis and greatly helps physicians to make decisions, it is an enormous opportunity to healthcare facility. Therefore, healthcare facilities should deploy and use ANFIS, and other AI methods, for diagnosing diseases. This avoids the problems related to traditional diagnosis methods and the associated risks.

As the conclusion, the following results can be summarized:

- It was seen that adaptive neuro-fuzzy inference system could be successfully used to help diagnosis of TB disease.
- Either of the tree diagnosis variables (smear, x-ray and ERP) can be used to diagnosis TB diseases if one of the variables shows true positive result.
- Only one or two of the tree diagnosis variables cannot accurately diagnosis TB diseases if these variables show false negative results.

## REFERENCES

1. JunHua Li, Lesley Pek Wee, Pradeep Ray Subhagata Chattopadhyaya, “E-Health Readiness Framework from Electronic Health Records Perspective,” The University of New South Wales, Sydney NSW 2052, Australia, 2010.
2. D. Kaseje. “Health Care in Africa: Challenges, Opportunities and an Emerging Model For Improvement,” *The Woodrow Wilson International Center for Scholars, 2006*
3. A. Alwan. “Strengthening national health information systems: challenges and response,” *Eastern Mediterranean Health Journal*, 22(11), pp. 840-850, 2016.
4. I. Chudi P. “Healthcare problems in developing countries,” *Medical Practice and Reviews,* 1(1), pp. 9-11, 2010.
5. A. Ikonje. “Strengthening data management and use in decision making to improve health care services: Lessons learnt,” *Dissemination Workshop Report*. Published by the Tibu Homa Project for the United States Agency for International Development. Mwanza, Tanzania: University Research Co., LLC, 2014.
6. O. Norheim F. et al. “Avoiding 40% of the premature deaths in each country, 2010–30: review of national mortality trends to help quantify the UN Sustainable Development Goal for health,” *Open Access article distributed under the terms of CC BY-NC-SA*, pp. 1-14, 2014.

7. WHO. “Global status report on noncommunicable diseases 2010,” *World Health Organization* Printed in Italy, 2011.
8. E. Petersen, A. Denmark. “Commemorating World Tuberculosis Day 2015”; *International Journal of Infectious Diseases*,” 32, pp. 1–4, 2015.
9. P. Dande, P. Samant. “Acquaintance to Artificial Neural Networks and use of artificial intelligence as a diagnostic tool for tuberculosis: A review; Tuberculosis 108 (2018) 1e9.
10. A. Zumla, E. Petersen, T. Nyirenda, J. Chakaya. “Tackling the Tuberculosis Epidemic in sub-Saharan Africa – unique opportunities arising from the second European Developing Countries Clinical Trials Partnership (EDCTP) programme 2015-2024,” *International Journal of Infectious Diseases*, 32 pp. 46–49, 2015.
11. WHO. “Global Tuberculosis Report $20^{th}$ edition,” 2015.
12. WHO. “Global Tuberculosis Report,” 2016
13. N. Walia, S. Kumar, R. Malhotra. “Design and Identification of Tuberculosis using Fuzzy Based Decision Support System,” *Advances in Computer Science and Information Technology (ACSIT),* 2 (8), pp. 57-62, 2015.
14. M. Omisorea O., O. Samuel W., E. Atajeromavwo J. “A Genetic-Neuro-Fuzzy inferential model for diagnosis of tuberculosis,” *Applied Computing and Informatics*, 13, pp. 27–37, 2017.
15. S. Kara, S. Dasb, P. Ghosh K. “Applications of neuro fuzzy systems: A brief review and future outline,” *Applied Soft Computing*, 15, pp. 243–259, 2014.
16. Y. Peng, X. Zhang. “Integrative data mining in systems biology: from text to network mining, *Artif. Intell. Med.,* 41 (2), pp. 83–86, 2007.
17. X. Zhou, B. Liu, Z. Wu, Y. Feng. “Integrative mining of traditional Chinese medicine literature and medline for functional gene networks, *Artif. Intell. Med*., 41 (2), pp. 87–104, 2007.
18. H. Sol G. “Expert systems and artificial intelligence in decision support systems, in: *Proceedings of the Second Mini-Euro Conference, 17–20 November 1985,* Springer, Lunteren, The Netherlands, pp. 1–2, 1987.
19. R. Wael, A. Ibrahim, M. Morcos M. “Artificial Intelligence and Advanced Mathematical Tools for Power Quality Applications: A Survey,” *IEEE transactions on power delivery,* 17(2), 2002.
20. Q. Al-Shayea. “Artificial neural networks in medical diagnosis, *Int J Comp Sci.Appl.*, 8(151), 2011.
21. A. Barwad, P. Dey, S. Susheilia. “Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology, Cytomet. B Clyn Cytom. 82, pp. 107–111, 2012.
22. T. Manikandan, N. Bharathi. “Hybrid neuro-fuzzy system for prediction of stages of lung cancer based on the observed symptom values,” *Biomedical Research*, 28 (2), pp. 588-593, 2017.
23. M. Alizadeh, M. Gharakhani, E. Fotoohi, R. Rada. “Design and analysis of experiments in ANFIS modeling for stock price prediction,” *International Journal of Industrial Engineering Computations*, 2, 409–418, 2011.
24. J. JSR. “ANFIS: adaptive network-based fuzzy inference systems. *IEEE Trans Sys Man Cybern*, 23, pp.665–685, 1993.
25. M. Vose D. “The Simple Genetic Algorithm. Cambridge, U.K.: MIT Press, 1999.
26. W. Suparta, K. Alhasa M. “Modeling of Tropospheric Delays Using ANFIS, Springer Briefs in Meteorology, pp.5-18, 2015.
27. A. Abraham. "Adaptation of Fuzzy Inference System Using Neural Learning", in Nedjah, Nadia; de Macedo Mourelle, Luiza, Fuzzy Systems Engineering: Theory and Practice, Studies in Fuzziness and Soft Computing, 181, Germany: Springer Verlag, pp. 53–83, 2005.
28. J. Mizutani S. “Neuro-Fuzzy and Soft Computing – Prentice Hall,”, pp 335–368, 1997.
29. P. Tahmasebi. "A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation" (PDF). *Computers & Geosciences*, 42, pp. 18–27, 2012.
30. T. Ucar, A. Karahoca, D. Karahoca. “Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets; *Neural Comput & Applic*, 23, pp. 471–483, 2013.
31. C. Go ́mez, R. Hornero, D. Aba ́solo, A. Ferna ́ndez, J. Escudero. “Analysis of MEG background activity in alzheimer’s disease using nonlinear methods and ANFIS. *Ann Biomed Eng*, 37(3), pp. 586–594, 2009.
32. J.S. Jang. “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), pp. 469–483, 1993.
33. Mohamed Taha, Hamed Nassar, Tarek Gharib and Ajith Abraham, An Efficient Algorithm for Incremental Mining of Temporal Association Rules, Data and Knowledge Engineering, Elsevier Science, Netherlands, 69:800-815, 2010.
34. Swagatam Das, Sambarta Dasgupta , Arijit Biswas, Ajith Abraham and Amit Konar, On Stability of the Chemotactic Dynamics in Bacterial Foraging Optimization Algorithm, IEEE Transactions on Systems Man and Cybernetics - Part A, IEEE Press, USA, 39(3): 670-679, 2009.
35. Tibebe Tesema, Ajith Abraham and Crina Grosan, Rule Mining and Classification of Road Accidents Using Adaptive Regression Trees, International Journal of Simulation Systems, Science & Technology, UK, 6(10-11):80-94, 2005.
36. Hongbo Liu, Ajith Abraham and Maurice Clerc, Chaotic Dynamic Characteristics in Swarm Intelligence, Applied Soft Computing Journal, Elsevier Science, 7(3):1019-1026, 2007.
37. Lizhi Peng, Bo Yang, Yuehui Chen and Ajith Abraham, Data Gravitation Based Classification, Information Sciences, Elsevier Science, Netherlands, 179(6):809-819, 2009.
38. Hesam Izakian, Behrouz Ladani, Kamran Zamanifar and Ajith Abraham, A Particle Swarm Optimization Approach for Grid Job Scheduling, Third International Conference on Information Systems, Technology and Management (ICISTM-09), Communications in Computer and Information Science, Springer Verlag, Germany, ISBN 978-3-642-00404-9, pp. 100-109, 2009.