International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 13308-13312

Genetic-Neuro-Fuzzy Inferential Model for Tuberculosis Detection

R. Beaulah Vathana[1,*], Dr. R. BalaSubramanian[2]

[1] Department of Computer Science and Engineering, Mepco Schlenk Engineering Sivakasi, India.
[2]Department of Computer Science Engineering, Manonmaniam Sundaranar University, Tirunelveli, India.
*(*Corresponding Author: R. Beaulah Jeyavathana)*

Abstract

Tuberculosis is one of the dangerous infectious diseases that can be characterized by the growth of tubercles in the tissues. This disease mainly affects the lungs and also the other parts of our body. The orthodox diagnosis methods available for Tuberculosis diagnosis were been faced with a number of challenges which can, if measure not taken, increase the spread rate; hence, there is a need for aid in diagnosis of the disease. This study proposes a technique for intelligent diagnosis of TB using Genetic-Neuro-Fuzzy Inferential method to provide a decision support platform that can assist medical practitioner in administering accurate, timely, and cost effective diagnosis of Tuberculosis. The medical record of 100 TB patients aged 15 to 75 were used to evaluate the performance of the multi-technique decision support system. 70% of the dataset was used training data, 15% was used for validation while the remaining 15% was used to observe the performance of the proposed system.

Keywords: Tuberculosis, Fuzzy Logic approach, Neural Network, Genetic Algorithm, Decision Support system.

INTRODUCTION

Tuberculosis is the infectious bacterial disease that is caused by the organism called Mycobacterium tuberculosis and that may affect any tissues of the body but it mainly affects the lungs. The TB is one of airborne pathogen and that can spread through air or by coughing or sneezing from one person to another. TB affects all age groups in all parts of the world. Mostly it affects young adults and also the peoples who are all in the developing countries. Active lung TB are cough with sputum and blood at a time, weight loss, chest pains, fever, weakness and night sweats. Tuberculosis bacteria are present in sputum samples are identified under a microscope. It detects only half the number of Tuberculosis cases and cannot detect drug-resistance.

In 2015, around 11 million people fell ill with TB and 2 million peoples were died from the disease. Over 95% of the deaths in TB occur in low and middle-income countries. Around Six countries account for 60% of the total, with India leading the count, followed by Pakistan, China, Indonesia, Nigeria and South Africa. An estimated one million children became ill with TB and 170000 children died of TB (excluding children with HIV).

The World Health Organization (WHO) estimated in 2006, that each year, more than 8 million new cases of TB occur and approximately 3 million persons die from the disease and estimated that between 19% and 43% of the world's population will be infected with Mycobacterium Tuberculosis. Within the past decade it has become clear that the spread of HIV infection and the immigration of persons from areas of high incidence have resulted in increased numbers of TB cases. It has always occurred disproportionately among disadvantaged populations such as the homeless, malnourished, and overcrowded. Today, several methods for the diagnosis of TB have been proposed. Tuberculin Test, Radiological Examination, and Sputum Smear Microscopy are common conventional approaches however in the last 10 years, several molecular methods have been developed for direct detection, identification and susceptibility testing of mycobacteria .

Orthodox methods of diagnosing TB are primarily through physical examination and laboratory tests. The former involves asking patients certain questions for prognosis purposes while tests are carried out to affirm physical examination. Diagnosis can be stopped if medical practitioner is totally convinced after physical examination however, this is not advised. This orthodox method is currently faced with a number of challenges such as lack of medical facilities in most medical centers and as a result, inhibiting the management of

TB in developing countries Orthodox methods of diagnosing TB are primarily through physical examination and laboratory tests. The former involves asking patients certain questions for prognosis purposes while tests are carried out to affirm physical examination. Diagnosis can be stopped if medical practitioner is totally convinced after physical examination however, this is not advised. This orthodox method is currently faced with a number of challenges such as lack of medical facilities in most medical centers and as a result, inhibiting the management of TB in developing countries.

The strength of IT in providing an effective and efficient solution to real life problems has been explored to aid scientific discoveries and advancement of different fields of medicine. Hence, to reduce the morbidity and mortality rates in human as a result of TB, there is need to incorporate IT into its diagnostic approach. This study, therefore, proposes a decision support model for intelligent diagnosis of TB using Genetic-Neuro-Fuzzy Inferential technique. The model is aimed at providing a decision support platform that can aid medical practitioners in administering accurate, timely, and cost effective diagnosis of TB in developing countries.

RELATED WORK

A novel approach to identify tuberculosis bacteria based on shape and colour was proposed by **M.Forero et al. (2004).**

Designed algorithm technique was based on combined use of invariant shape features together of bacilli with simple thresholding operation on chromatic channels. This methodology is based on segmentation followed by an identification procedure, for which 110 samples of bacilli was analyzed. Usefulness of K-means clustering algorithm techniques was applied to predict classification, accuracy, and sensitivity versus specificity was evaluated using ROC analysis procedure. Further, the author suggested exploring a colour-based edge segmentation technique using derivative operators to all chromatic channels and by using Bayesian decision theory [1].

N. Walia et al. (2015) had presented a systematic approach for design and identification of tuberculosis using fuzzy based decision support system. Their framework briefly explains relation between different input attributes and its symptoms. Author concluded that fuzzy basis dependent expert systems can be used during diagnosis. Further,author suggested that designed system can be extended for construction of other chronic obstructive diseases using hybrid neuro systems [2].

An integrated approach for automated detection of early lung cancer and tuberculosis based X-ray image analysis was demonstrated by **K. Lee (2006).** Various symptoms of the disease and finding nodules were focused during this paper. The proposed technique uses watershed segmentation approach to isolate a lung X-ray image, and then apply a small scanning window to determine whether any pixel is a part of a disease nodule or not. Additionally, various methods used to detect early signs of cancer and tuberculosis was also explained in this work [3].

Wenbo Li, Yan Kang (2015) using a new adaptive VOI selection method. Twenty-two features were extracted to distinguish nodules, vascular endpoint or vascular cross structure, designed an optimal feature combination selection frame based on improved genetic algorithm and support vector machine. The improved GA algorithm is used to select the optimal feature combination from the feature pool to establish SVM classifier [4].

Asogbon MG(2016), proposed enhanced Neuro-Fuzzy system based on genetic algorithm for medical diagnosis proposed the use of Genetic Algorithm (GA) technique to automatically evolve optimum connection weights needed to efficiently train a built ANFIS model used for Typhoid fever diagnosis. The GA module computes the best set of connection weights, stores them, and later supplies them to the corresponding hidden layer nodes for training the ANFIS[5].

An artificial intelligent approach for estimation of disease and resource utilization was discussed by **E.Papageorgiou et al. (2009).** Fuzzy cognitive map based tool was used to represent medical diagnosis system concentrated on relating to lungs infections. Due to easy graphical representation approach, the proposed method makes wide use of computer consultation system. The presented system would offer a solution for requirements imposed by the target application, disease symptoms, signs and laboratory tests [6].

Usefulness of fuzzy logic approach to decision support system in medicine was discussed by **U. Dev et al. (2011**). This approach was based on the diagnosis of a patient suffering heart failure treated with beta blockers. The developed system is a prototype warning system for clinical problems which is based on the assumption that can be analyzed using simple rules. The planned technique generates basic rules using fuzzy logic based on expert experience [7].

A spectrum of soft computing decision-making model to solve a real life complex problem related with medical science was explored by **P.Srivasta et al. (2013)**. The designed network was tested with ECG analysis and the satisfactory factor was measured under a domain of considered inputs[8].

A novel Intuitionistic fuzzy cognitive map (iFCM) based on theory of Intuitionistic fuzzy sets was explained was **M. Arts et al. (2013**). This model offers checking and classification techniques to predict human decision model. The proposed system has an extension of FCM to the co-evaluate degree of hesitation; experts may suffer while defining a relation between concepts of FCM. The author demonstrates the effectiveness of FCM with numeric reproducible expels on a process of control and decision support. The simulation studies describe the performance of iFCM for medical decision support platform and the results obtained were significantly better than obtained with conventional FCM model. In this work, fuzzy rules are applied to determine the stage of tuberculosis

SYSTEM DESIGN

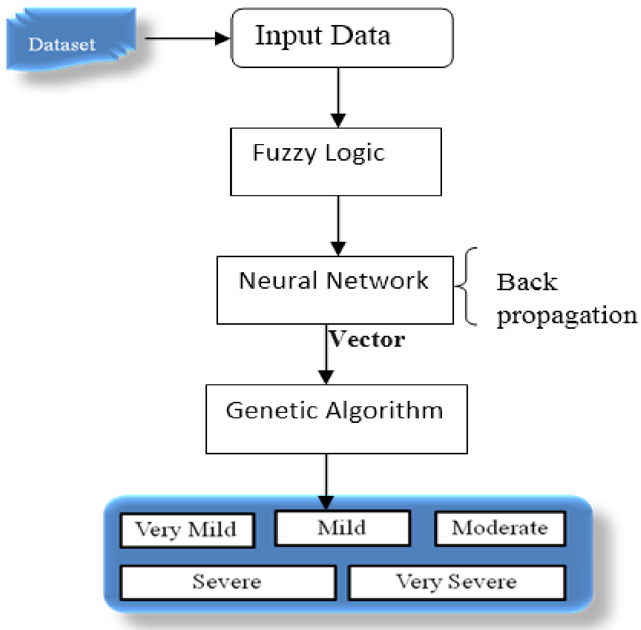


Figure 1. System Design

Fig.1, shows the system design of the proposed method. This is an inferential technique proposed to integrate GA, NN and FL components to provide a self-learning and adaptive system for handling uncertain and imprecise data for diagnosis of tuberculosis. The inference engine consists of reasoning algorithm driven by the production rules based on Mamdani's Inference Mechanism.

PROPOSED SYSTEM

MATERIAL AND METHODS

In this system, dataset are collected for 100 TB patients from the goverment hospital of tirunelveli. The medical record of 100 TB patients aged 15 to 75 were used to evaluate the performance of the multi-technique decision support system. 70% of the dataset was used training data, 15% was used for validation while the remaining 15% was used to observe the performance of the proposed system.

DATABASE

Structured database presents quantitative data about facts and the established rules in the field of medicine focusing on diagnosis of TB. The facts comprise of signs and symptoms of TB, while rules are patterns to draw deductions based on available information . Unstructured database is heuristic in nature and hence gathered by experience, good practices, guesses, and judgments . The database comprises of Patient-Bio-Data, Disease-Physical-Signs, Disease-Symptoms, Medical-History, Physical Examination, results of diagnostic tests and Patient Diagnosis.

FUZZY LOGIC

The diagnosis process harnesses the strength of fuzzy logic component in the following operational sequence:

Fuzzification of input variables:

Fuzzification is a process that determines the degree of membership to the fuzzy set based on fuzzy membership function. The first step is to

- Create a fuzzy set of the parameters. The parameters will be described with five linguistic variables (very mild ,mild, moderate, severe and very severe).
- The degree of membership for a fuzzy system is of the range [0 1]. A range of the fuzzy value using the linguistic variables will be determined by the expert.

  e.g. mild $0.1 \le x \le 0.3$.

Establishment of fuzzy rule base:

The rule base for TB diagnosis is characterized by a set of IF–THEN rules in which the antecedents (IF parts) and consequents (THEN parts) involve linguistic variables. The rules can be formulated with assistance of experts in the management of TB, or on consultation to existing standard literature. A rule can only fire if any of its precedence parameters such as very mild, mild, moderate, severe, and very severe evaluates to TRUE, otherwise it does not fire.

Fuzzy Inference:

The inference engine controls how the rules are applied towards facts. This is the part of rule-based expert system that makes inferences. It decides which rules are satisfied by facts and controls overall execution. Also, it matches the facts against the rules to see what rules are applicable. The system will make use of forward chaining reasoning; it would make use of the facts given by the patient to diagnose the problem. Fuzzy inference is the process of mapping from a given input to an output using the theory of fuzzy sets. Rules are used in the knowledge-base by the fuzzy inference engine to derive conclusion based on the rules.

Defuzzification:

This involves changing fuzzy output back into numerical values for system action. The output from the inference engine is translated into crisp output which is more precise than the fuzzy output.

NEURAL NETWORK

Neural Network has the capability of capturing domain knowledge from available indicators and can readily handle both continuous and discrete data. NN is used to train and test the designed fuzzy system to optimize the performance of the overall system. Each diagnosis variable has a weight W_i which shows its contribution in the diagnosis process.

The raw information obtained from patients is fed into NN via input layer and participation of each category of variables is determined at a hidden layer of the network using:

$$CAT_i = \sum_{i}^{n} A_i * W_{Ai}$$

CATi is ith category of variable, n is count of variables in CATi, and Ai is the ith diagnosis variable with weight W_{Ai}

Result of the output layer represents an overall output of diagnosis by the NN component. The output result is given by,

$$Output_{NN} = \sum_{i}^{n} CAT_i * W_{CATi}$$

where W_{CATi} is the connection weight of CAT_i

GENETIC ALGORITHM

Actually, NN provides a structure for combining the diagnostic parameters which could serve as a platform for the inference engine, but a specific issue with NN is lack of definite way of determining the connection weights for hidden layers when dealing with a particular problem.

A number of medical diagnosis had been assisted by neuro-fuzzy systems though such systems had been built based on trial and errors, this increases computation cost. In this study, genetic optimization is performed to choose optimal values from a group of diagnostic parameters which serve as input. There are 23 diagnostic parameters in the NN but the task is to

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 13308-13312

decide which parameters are taken as input in order to minimize complexity.

An individual chromosome consists of 24 genes and each gene represents the connection weight of a diagnosis variable in a length of 1 bit.

One feasible solution is to generate an initial population holding a set of possible solutions from random chromosomes. A chromosome is represented as a vector C = (C_A,…,C_X) of binary decision variables C_i=0,2,3 encoded in binary representation as string consisting {0, 1} genes.

A gene C_i =1 if the ith variable is included in a solution set of a diagnostic process otherwise 0. Fitness function is used to optimize each chromosome by evaluating the genes that constitute the chromosome using their fitness value.

As evolutionary algorithm continues through its cycle, fitness value of each chromosome keeps improving till it reaches an optimum value when it can no longer improve.

Finally, it shows chromosomes of some candidates and their fitness values. A number of constraints have been considered in carrying out appropriate management of disease in medical diagnosis, therefore fitness evaluation of chromosome must be done with proper constraint validation.

Constraints can be termed as objectives that must be achieved in which some render most of the solutions from the search space. The fitness evaluation of an individual F(i) is done as:

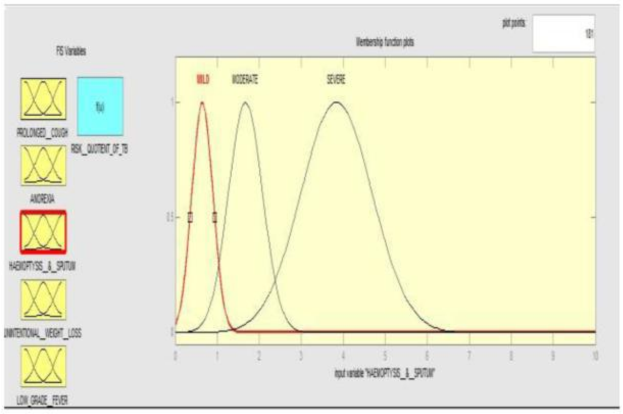
$$F_i = (1 + \sum_{i=0}^{n} W_i * C_i(p))^{-1}$$

where n is the number of diagnosis variables, W_i is the weight associated with ith variable and $C_i(p)$ is the number of violations for ith constraint at solution p.

This fitness function has a range of [0, 1] and an optimal solution occurs when we have 0 violations thus $\sum_{i=0}^{n} W_i * C_i(p)$ which results in F_i=1. Chromosomes with higher fitness values are selected as parents for mating in order to produce outstanding candidates and maximize the fitness function. The probability of choosing an individual for genetic operation is proportional to its fitness. This process is repeated until an optimal connection weight is achieved.

SIMULATION BASED RESULTS AND DISCUSSIONS

Membership function plot: The potential symptoms are taken as input and membership functions are defined according to the level of severity of the disease.



Performance Analysis

Given a True Positive value (TP) that represents the number of patients with tuberculosis as agreed by both model and human expert, a True Negative value (TN) indicating the number of patients where agreement could not be reached by both model and human expert, and TNR as the total number of records; the sensitivity and accuracy of this model are:

Sensitivity= TP/TNR * 100%

Accuracy= TP+TN/TNR * 100%

Rule Viewer for Pulmonary TB : On the basis of the if-then rules defined, a set of values are obtained and processed using rule viewer. The Rule Viewer presents the on sight view of the fuzzy inference systems process. The Rule Viewer also depicts how the shape of certain membership functions influences the final result. Each rule is a row of plots, and each column is a variable. The system output obtained using weighted average defuzzification process. All output membership functions should be of similar type whether linear or a constant.

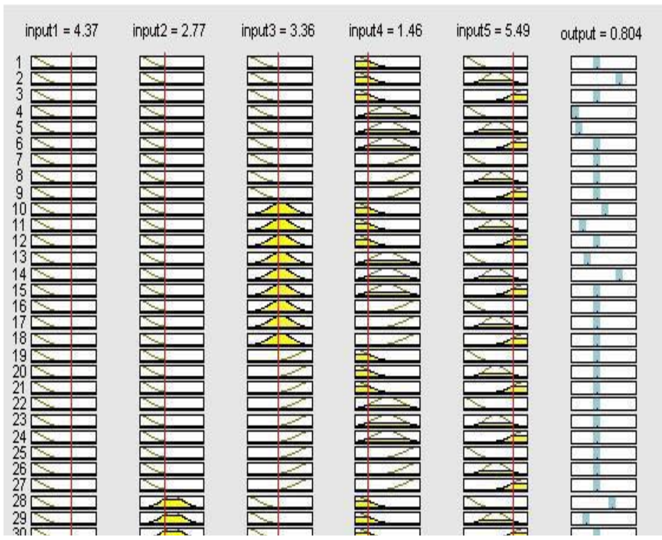


Table 1. Parameters used in finding optimal solution

S.No	Parameters	Value
1	Number of Generations	20
2	Number of individuals	40
3	Crossover probability	0.55
4	Mutation probability	0.35

In this simulation, the GA used selected values given in Table1 to find optimal solution and once a complete set of optimal value for the parameters is reached, the GA processes stops. High mutation probability used in the parameter settings is a brain behind bringing new individuals at each stage hence, to avoid a scenario whereby best combination will only be from initial individuals.

The sensitivity and accuracy of the proposed model are 72% and 82%.

CONCLUSION

The use of soft computing techniques in medical diagnosis cannot be overemphasized as they have greatly imparted the processes in medical diagnosis and aided an increase in diagnosis accuracy. This work, demonstrated how an aggregation of such technique can assist in the diagnosis of TB.

REFERENCES

[1] C.A. Robert, J.E. Buikstra, The Bio-Archaeology of Tuberculosis: A Global Perspective on a Re-Emerging Disease, University Press of Florida, Gainesville, Fl, 2009.

[2] David M. Morens, Gregory K. Folkers, Anthony S. Fauci, The challenge of emerging and re-emerging infectious diseases, Nature 430 (2004) 242–249.

[3] World Health Organization, WHO Country Cooperation Strategy: Federal Republic, 2002.

[4] World Health Organization, Global Tuberculosis Control, WHO Report 2008, Geneva, 2008.

[5] P. Sudre, G.T. Dam, A. Kochi, Tuberculosis: a global overview of the situation today, Bull. World Health Organ. 70 (1992) 149– 159.

[6] American Thoracic Society, Diagnostic standards and classification of tuberculosis in adults and children, Am. J. Respir. Crit. Care. Med. 161 (2000) 1376–1395.

[7] Syed Tasleem, Gulraiz Ali, Farzana Mahdi SikandarHasan, Molecular diagnosis of tuberculosis: a new primer design, Iran. J. Clin. Infect. Dis. 4 (2) (2009) 105–108.

[8] Young S. Kim, Donald E. Shelton, Gregg Barak, Studying juror expectations for scientific evidence: a new model for looking at the CSI myth, J. Am. Judges Assoc. 47 (1–2) (2011) 8–18.

[9] S.A. Oke, A literature review on artificial intelligence, Int. J. Inf. Manage. Sci. 19 (4) (2008) 535–570.

[10] V.P. Masnikosa, The fundamental problem of an artificial intelligence realisation, Cybernetes 27 (1) (1998).

[11] W.E. Halal, Artificial intelligence is almost here, On the Horizon, The Strategic Planning Resource for Education Professionals, vol. 11(2), 2003.

[12] Y. Peng, X. Zhang, Integrative data mining in systems biology: from text to network mining, Artif. Intell. Med. 41 (2) (2007) 83– 86.

[13] X. Zhou, B. Liu, Z. Wu, Y. Feng, Integrative mining of traditional chinese medicine literature and medline for