ORIGINAL PAPER

Tuberculosis Disease Diagnosis Using Artificial Neural Network Trained with Genetic Algorithm

Erhan Elveren · Nejat Yumuşak

Abstract Tuberculosis is a common and often deadly infectious disease caused by mycobacterium; in humans it is mainly *Mycobacterium tuberculosis* (Wikipedia 2009). It is a great problem for most developing countries because of the low diagnosis and treatment opportunities. Tuberculosis has the highest mortality level among the diseases caused by a single type of microorganism. Thus, tuberculosis is a great health concern all over the world, and in Turkey as well. This article presents a study on tuberculosis diagnosis, carried out with the help of multilayer neural networks (MLNNs). For this purpose, an MLNN with two hidden layers and a genetic algorithm for training algorithm has been used. The tuberculosis dataset was taken from a state hospital's database, based on patient's epicrisis reports.

Keywords Diagnosing tuberculosis ·
Multilayer neural network · Genetic algorithm

Introduction

Tuberculosis is an infectious disease mostly caused by the microorganism called *Mycobacterium tuberculosis* [1]. These microorganisms are generally brought in together with the infected air. Blood and lymph are the common means for them to 'transport' from the initial spot in the lungs to all other parts of the body; the infection can also set in through the airways and/or in case of immediate exposure of other organs. When an infected person coughs, sneezes, talks or spits, the bacteria get into the air. Quite a small quantity of bacteria is enough to get infected. The risk of being infected depends on how long or how intense the exposure to the bacteria is. The risk is greatest in those who live or work with an infected person [2–4].

The typical symptoms of pulmonary tuberculosis are the chronic cough, weight loss, intermittent fever, loss of appetite, night sweats and coughing out blood. The disease develops in the human body in two stages. At the first stage, the infection sets in, having been 'imported' from an infected individual (this stage is commonly referred to as "tuberculosis infection"); consequently, at the second stage, the disease develops ("tuberculosis" itself) [2, 3].

Tuberculosis is one of the commonest major maladies worldwide, especially in Asia and Africa. It is a great problem for most developing countries. It has the largest mortality ratio for an illness caused by a single type of microorganism. In 2006, 9.2 million new cases of infection and 1.7 million deaths were reported to have occurred throughout the world. Of them 0.7 million cases and 0.2 deaths were HIV-positive people [2, 4]. In Turkey, according to the State's Health Ministry's 2006 report, approximately 23,875 patients died of tuberculosis [4].

The multilayer neural networks (MLNNs) have been successfully implemented in decision support systems for disease diagnosis systems [5–8]. The back-propagation (BP) algorithm [9] is widely used for training the network. However, it has to take too many steps to train the network, and the weights are calculated step by step. In genetic algorithm, several weight sets are calculated in parallel and thus, finding the optimal solution needs less complicated computations.

This paper aims to present a study on diagnosing tuberculosis with the help of multilayer neural network

E. Elveren (✉) · N. Yumuşak
Department of Computer Engineering, Sakarya University,
54187 Adapazari, Turkey
e-mail: erhanelveren@duzce.edu.tr
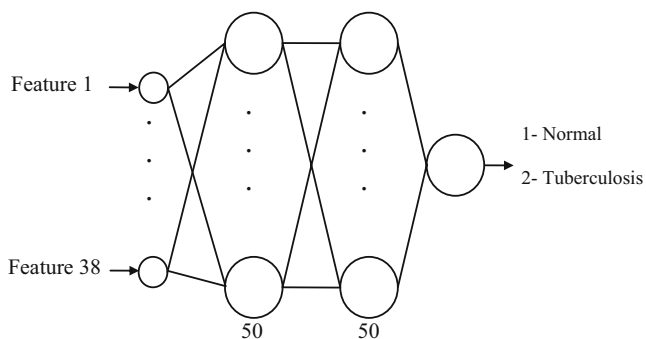
N. Yumuşak
e-mail: nyumusak@sakarya.edu.tr

Fig. 1 The implementation of multilayer neural network used for diagnosing tuberculosis

trained with genetic algorithm. For this purpose, an MLNN with two hidden layers and genetic algorithm has been used (Fig. 1). The results were also compared with the previous studies [10–12]. The tuberculosis dataset was taken from a state hospital's database, based on patient's epicrisis reports.

Data source

In order to perform this study, first, the patients' epicrisis reports have been taken from Diyarbakir Chest Disease Hospital. The dataset was prepared, based on these epicrisis reports, which contained 150 samples for two classes of patients: those with tuberculosis (Class 1: 50 patients), and those with other, non-tuberculosis diseases (Class 2: 100 patients).
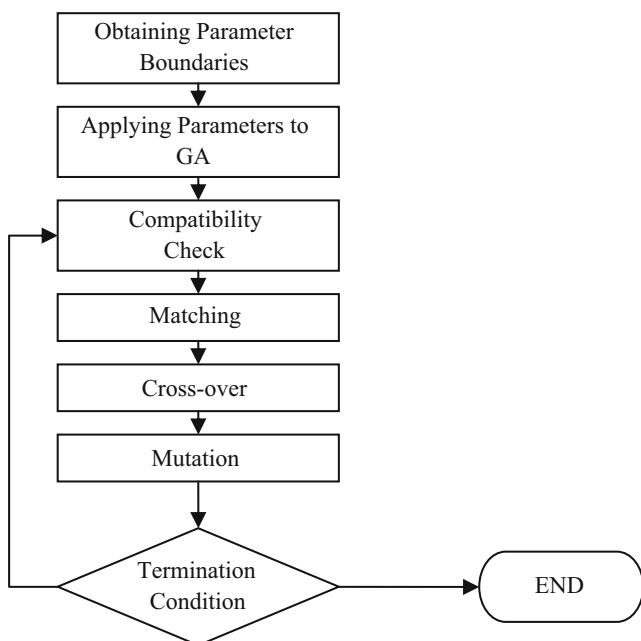

Fig. 2 The flow-chart of the simple genetic algorithm

All samples present 38 parametres. Organized into 2 groups, these parametres are:

1. *The results of preliminary examination*: complaints of cough, high temperature, aches in the chest, weakness, dyspnea, wheezes in the chest, sputum, sound on respiratory tract, pressure in chest, smoking

2. *Laboratory tests*: leukocyte (WBC), erythrocyte (RBC), trombosit (PLT), hematocrit (HCT), hemoglobin (HGB), albumin 2, alkaline phosphatase 2°L, alanine aminotransferase (ALT), amylase, aspartate aminotransferase (AST), bilirubin (total+ direct), CK/ creatine kinase total, CK-MB, iron (SERUM), gamma-glutamyl transferase (GGT), glucose, HDL cholesterol, calcium (CA), blood urea nitrogen (BUN), chlorine (CL), cholesterol, creatinin, lactic dehydrogenase (LDH), potassium (K), sodium (NA), total protein, triglyceride, uric acid.

Previous studies

There have been several studies on diagnosing tuberculosis [10–12]. El-Solh, et al., used a general regression neural network (GRNN) and clinical and radiographic information as parameters to predict active pulmonary tuberculosis [10]. In that study, 21 distinct parameters, which fall into three groups: demographic variables, constitutional symptoms and radiographic findings, have been used. The authors utilized a ten-fold cross-validation procedure to train the network and provided estimation for diagnosing tuberculosis. The authors reported approximately 92.3% diagnosis accuracy [10]. Er, et al., used two different multilayer neural networks with one and two hidden layers and trained the network with Levenberg-Marquardt training algorithm [12]. The input patterns were formed by the 38 parameters, which have been implemented in this study as well. Three-fold cross-validation procedure was utilized to train the network. The authors reported approximately 93.93% diagnosis accuracy, using back-propagation algorithm with momentum (BPwM) [12]. Santos, et al., used symptoms and physical signs for constructing the neural network modeling [11].

Table 1 Compared studies

Study	Method	Classification accuracy (%)
[10]	GRNN (one hidden layer)	92.30%
[11]	MLNN with BP (one hidden layer)	77.00%
[12]	MLNN with BPwM (two hidden layers)	93.93%
This study	MLNN with GA	94.88%

They reported approximately 77% diagnosis accuracy using a MLNN with one hidden layer.

Diagnosing tuberculosis with the help of neural networks

In this study, a multilayer neural network structure with two hidden layers is used for diagnosing tuberculosis.

There are 50 neurons on each of the hidden layers and only one neuron at the output layer. All of the neurons use a non-linear sigmoid activation function. In this system, 38 inputs were understood as features and one output represents the value of the diagnosis (1—other, 2—tuberculosis). To train the MLNN, genetic algorithm was used. Detailed information can be found in the references [13].

If a neural network learns the training set of a task, it generalizes the problem and gives more or less accurate outputs for the dataset different from training set. But if a neural network starts to memorize the training set, it gives proper answers to the training set but the generalization performance falls down. The k-fold cross-validation method shows how accurate a generalization can be made by a neural network [17]. In this study, three-fold cross-validation method is used to estimate the performance of the neural network used.

As a performance measure, the formula below is used [6]:

$$classification\ accuracy = \frac{\sum_{i=1}^{|N|} assess(n_i)}{|N|}\ n_i \in N$$

$$assess(n) = \begin{cases} 1\ if\ classify(n) \\ 0\ \ \ otherwise \end{cases}$$

where N is the set of data to be classified (validation set), nc is the class of the item n, and $classify\ (n)$ returns the classification of n calculated with neural network.

Training the network with genetic algorithm

Genetic Algorithm is a search technique to find approximate solutions to optimization problems. It is a global search technique and a particular class of evolutionary algorithms. From biological sciences, evolutionary processes have been borrowed and translated to efficient search and design strategies. Genetic Algorithms use these strategies to find an optimum solution for any multi-dimensional problem [14].

When the genetic algorithm is used for training an Artificial Neural Network (ANN), weights and biases are encoded into chromosomes. So a generation (a set of chromosomes) represents a set of ANN.

Each chromosome is decoded and the error function of ANN, which the chromosome represents, is calculated. In order to train the ANN, the error over all training sets has to be minimized [15–18] (Fig. 2). With a chosen selection ratio, more successful chromosomes are selected to create new generation by cross-breeding with those, being used to create more successful chromosomes than their predecessors. After this operation, within a mutation ratio, chromosomes are mutated to obtain diversity with possible successful generations. The mutation helps to change the improper chromosomes but can as well damage the proper ones. That's why the mutation ratio must be chosen carefully.

Results and conclusion

The classification accuracies obtained by other studies for tuberculosis are; by reference [10] was 92.30% and [12] was 93.93% with back-propagation. More efficiency was obtained by using Levenberg-Marquardt training algorithm [12], but because of the complexity of the algorithm, it is more difficult to encode it. For that reason, the results were compared with back-propagation algorithm. In this study, the result was 94.88% with the same structure as in the reference [12].

As seen in the Table 1, this study shows that the general algorithm as a training algorithm has better performance in diagnosing tuberculosis diagnosis using multilayer neural network with simple coding.

By way of conclusion, the following results should be mentioned:

- In general, the classification success obtained by this study is better than those obtained by reference studies [10–12];
- Training the multilayer neural network with genetic algorithm takes too much time;
- It is obtained that using MLNN with GA is a successful way to diagnose tuberculosis;

References

1. Wikipedia, http://en.wikipedia.org/wiki/Mycobacterium_tuberculosis. Last accessed: 23 May, 2009.
2. Enarson, D. A., Rieder, H. L., Arnadottir, T., and Trébucq, A., A guide for low income countries. Paris: International Union Against Tuberculosis and Lung Disease, ISBN: 2-914365-00-4, 2000.
3. Royal College of Physicians of London, Tuberculosis. London: Royal College of Physicians of London, ISBN: 1 86016 277 0, 2006.
4. World Health Organization, Global tuberculosis control 2008 surveillance planning financing. ISBN: 978 92 4 156354 3, Geneva, 2008.
5. Kayaer, K., and Yıldırım, T., Medical diagnosis on Pima Indian diabetes using general regression neural networks. In Proc. of

International Conference on Artificial Neural Networks and Neural Information Processing (ICANN/ICONIP): Istanbul, (pp. 181–184), 2003.

6. Delen, D., Walker, G., and Kadam, A., Predicting breast cancer survivability: a comparison of three data mining methods. *Artif. Intell. Med.* 34:2113–2127, 2005.

7. Temurtas, F., A comparative study on thyroid disease diagnosis using neural networks. *Expert Syst. Appl.* 36:944–949, 2009. doi:10.1016/j.eswa.2007.10.010.

8. Er, O., and Temurtas, F., A study on chronic obstructive pulmonary disease diagnosis using multilayer neural networks. *J. Med. Syst.* 32:5429–5432, 2008. doi:10.1007/s10916-008-9148-6.

9. Rumelhart, D. E., Hinton, G. E., and Williams, R. J., Learning internal representations by error propagation. In: Rumelhart, D. E., and McClelland, J. L. (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1. Cambridge: MIT Press, pp. 318–362, 1986.

10. El-Solh, A. A., Hsiao, C.-B., Goodnough, S., Serghani, J., and Grant, B. J. B., Predicting active pulmonary tuberculosis using an artificial neural network. *Chest.* 116:968–973, 1999. doi:10.1378/chest.116.4.968.

11. dos Santos, A. M., Pereira, B. B., and de Seixas, J. M., Neural networks: an application for predicting smear negative pulmonary tuberculosis. Proceedings of the Statistics in the Health Sciences, March, 2004.

12. Er, O., Temurtas, F., and Tanrıkulu, A. Ç., Tuberculosis disease diagnosis using artificial neural networks. *J. Med. Syst.*, 2008. doi:10.1007/s10916-008-9241-x.

13. Montana, D. J., and Davis, L., Training feedforward neural networks using genetic algorithms. Cambridge: BBN Systems and Technologies Corp., 1989.

14. Goldberg, D. E., Genetic algorithms in search, optimization, and machine learning. London, Addison-Wesley, 1989.

15. Černohorský, D., Raida, Z., Škvor, Z., and Nováček, Z., *Analýza a optimalizace mikrovlnných struktur. (Analysis and optimization of the microwave tructures)*. VUTIUM: Brno, 1999. ISBN 80-214-1512-6.

16. Demuth, H., and Beale, M., Neural network toolbox for use with MATLAB. User's Guide (nnet.pdf in MATLAB help), 1993.

17. Raida, Z., Modeling EM structures in the neural network toolbox of MATLAB. *IEEE Antennas & Propagation Magazine*. 44 (6) 46–67, 2002. ISSN 1045-9243.

18. Gupta, M., M., Jin, L., and Homma, N., Static and dynamic neural networks from fundamentals. New York: Wiley, 2003.