Journal of Physics: Conference Series **1450** (2020) 012122 doi:10.1088/1742-6596/1450/1/012122

Design of pneumonia and pulmonary tuberculosis early detection system based on adaptive neuro fuzzy inference system

M Y Santoso[1], A M Disrinama[2], H N Amrullah[2]

[1] Department of Marine Engineering, Politeknik Perkapalan Negeri Surabaya, Jalan Teknik Kimia, Surabaya, Indonesia
[2] Department of Safety Engineering, Politeknik Perkapalan Negeri Surabaya, Jalan Teknik Kimia, Surabaya, Indonesia

Email: yusuf.santoso@ppns.ac.id

Abstract. The results of Basic Health Research in 2018 showed the prevalence of pneumonia and pulmonary tuberculosis (TB) in Indonesia 4.0 percent and 0.4 percent, respectively. However, with a minimal number of lung specialists, the handling of lung disease will be too late. There are only 600-700 lung specialists in Indonesia. This amount is very less when compared with existing lung disease cases. The use of ANFIS for early detection of lung disease is growing. However, the system designed is still used for one type of disease. This research will design an expert system based on ANFIS to detect lung disease early, i.e. for pneumonia and pulmonary TB. Subtractive clustering is used for clustering process. The results of the training showed that both models were able to give better performance compared to the model built using conventional clustering methods. The test results show that both models have comparable performance compared to their counterpart.

1. Introduction

The results of Basic Health Research in 2018 showed the prevalence of pneumonia and pulmonary tuberculosis (TB) in Indonesia 4.0% and 0.4%, respectively [1]. The high cases of lung disease in Indonesia makes this type of disease need serious attention. This is coupled with a lack of public awareness of lung health. In addition, air pollution is increasing due to smoke from active smokers, factory industrial fumes, motor vehicle fumes and various other pollution [2]. During this time, detection of diseases/pulmonary disorders are generally carried out examination of physical symptoms by doctors [3]. However, with a minimal number of lung specialists, the handling of lung disease will be too late. There are only 600-700 lung specialists in Indonesia [4]. This amount is very less when compared with existing lung disease cases.

The existence of an expert system is currently needed. This system can play a role to help the community in the absence of an expert [5]. Expert systems in health have been growing lately, one of which is the Adaptive Neuro Fuzzy Inference System (ANFIS) [6]. The expert system in health is not intended to replace the role of a doctor, but to help diagnose the patient's condition earlier [7]. Expert system as a tool to diagnose diseases with out of breath symptoms and also provide treatment solutions can be carried out through consultation by answering each a yes or no question [8]. Simulations of early detection of pulmonary disease based on patient symptoms by the Bayesian method [3] and X-ray image analysis based on image processing were successfully carried out [9].

The use of ANFIS for early detection of lung disease is growing. However, the system designed is still used for one type of disease. This research will design an expert system based on ANFIS to detect lung disease early, i.e. for pneumonia and pulmonary TB. If it has been successfully designed and tested, this system can later be turned on to health service sites in areas where the number of lung specialists is small. With this system, it is expected that more and more patients will be detected earlier about the illness, so that treatment will be made easier.

2. Methods

2.1. Pneumonia and Pulmonary TB Symptoms

Lung disease can be classified as infectious and non-communicable. Infectious pulmonary diseases include pneumonia and pulmonary TB [1]. Both types of disease have some of the same symptoms [10, 11]. However, there are some more specific symptoms that make handling differently for each disease. The symptoms for pneumonia and TB that are used as input for the expert system to be designed are shown in Table 1.

Table 1. The symptoms of pneumonia and pulmonary TB.

Lung Disease	Symptoms
Pneumonia	Cough Fever Headache Malaise Shivering
Pulmonary TB	Cough Fever Malaise Diaphoresis Shivering

2.2. Adaptive Neuro Fuzzy Inference System

Adaptive Neuro Fuzzy Inference System (ANFIS) is a fuzzy inference system (FIS) which illustrated in neural network architecture. It uses the first order Takagi-Sugeno-Kang (TSK) model, for computation simplicity and convenience [12]. The architecture consists of 5 layers, as shown in Figure 1. Each layer represents mathematical equations. The square node is an adaptive node. It means that parameter's value can change during training process. The circle node is non adaptive node. ANFIS input is represented by *x* and *y*, while the output is represented by *f*.

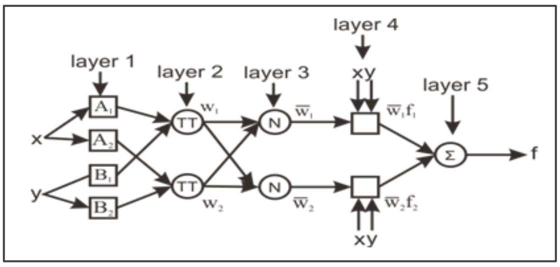

Figure 1. ANFIS structure [12].

Mathematical equation for Layer 1 depends on type of membership function (MF). Equation (1) and (2) show the function if gaussian MF is selected for the input. The function of Layer 2 is to multiply every input signal which comes from the previous layer output, as shown is Equation (3). Node number in this layer shows the created rule number.

$$O_{1,i} = \mu_{Ai}(x) = e^{\frac{-(x-c)^2}{2\sigma^2}} \; for\ i = 1,2 \tag{1}$$

$$O_{1,i} = \mu_{Bi}(y) = e^{\frac{-(y-c)^2}{2\sigma^2}} \; for\ i = 1,2 \tag{2}$$

$$O_{2,i} = w_i = \mu_{Ai}(x).\, \mu_{Bi}(x), \;\; i = 1,2 \tag{3}$$

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1+w_2}, \;\; for\ i = 1,2 \tag{4}$$

$$O_{4,i} = \overline{w_i}\, f_i = \overline{w_i}(\, p_i x + q_i y + r_i) \tag{5}$$

$$O_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{w_i} \tag{6}$$

Every node in Layer 3 is non adaptive node that show normalized firing strength of the node. In Layer 4, there are normalized firing strength from Layer 3 and parameter p, q, r, called consequent parameter. Layer 5 has a single node for summing all outputs form Layer 4. The output is a decision from the designed system.

This research design two ANFIS structures for pneumonia and pulmonary TB early detection system. Pneumonia system has seven inputs, i.e. cough, cough length, fever, fever length, headache, malaise, and shivering. For pulmonary TB, the inputs are cough, cough length, fever, fever length, malaise, diaphoresis, and shivering. The outputs for those systems either risk or not. The amount of data used to design pneumonia and pulmonary TB early detection systems is 76 patient data. Fifty-one data is used for the training process, while 25 for testing.

2.3. Subtractive Clustering

Subtractive clustering method was proposed by Chiu [13]. The method makes each data points are considered as the candidates for cluster centre. In subtractive clustering, a data point with the highest potential, which is a function of the distance measure, is considered as a cluster centre. The potential of each data point is estimated by the Equation (7).

$$P_i = \sum_{j=1}^{n} e^{-\alpha\|x_i - x_j\|^2} \tag{7}$$

where $\boldsymbol{\alpha} = \frac{\boldsymbol{\gamma}}{\boldsymbol{r_a^2}}$

Pi is the potential of i’th data point, n is the total number of data points, x_i and x_j are data vectors in data space including both input and output dimensions, γ is a positive constant and is selected as 4, and r_a is a positive constant defining the neighbourhood range of the cluster or simply the radius of hypersphere cluster in data space.

Subtractive clustering has four parameters, namely, accept ratio $\bar{\boldsymbol{\varepsilon}}$, reject ratio ε, cluster radius r_a and quash factor η (or r_b). These parameters have influence on the number of rules and error performance measures. Large values of $\bar{\boldsymbol{\varepsilon}}$ and ε will result in small number of rules. Conversely, small values of $\bar{\boldsymbol{\varepsilon}}$ and ε will increase the number of rules. A large value of ra generally results in fewer clusters that lead to a coarse model. A small value of ra can produce excessive number of rules that may result in an over-defined system.

For each designed ANFIS system, the parameters values are 0.5 for accept ratio, 0.15 for reject ratio and 1.25 for quash factor. Variation were made for cluster radius between 0 and 1 with increment 0.1. The selected model is model with smallest Root Mean Square Error (RMSE).

3. Results and discussion

The training data is used to design ANFIS with cluster radius variations in the clustering process. The model chosen for the pneumonia early detection system is a model with a cluster radius value of 0.3, because it has the smallest RMSE value. For the early detection system for pulmonary TB, the model chosen is the ANFIS model with a r_a value of 0.1. Table 2 shows the results of the RMSE calculation for each model of the early detection system. Each is then compared with a system built using the conventional two MF clustering method. Models built with subtractive clustering provide better performance. This result is because it generates the number of membership functions naturally according to system requirements [13].

The testing data is used to test the selected model. The RMSE results for the early detection system for pneumonia and pulmonary TB are shown in Table 3. The early detection system for pneumonia using subtractive gives better performance compared to models built using conventional clustering methods. For early pulmonary TB detection systems, both models have comparable performance.

Table 2. RMSE value for training data.

Clustering Methods	Pneumonia	Pulmonary TB
Subtractive	0.140	0.099
2 MF	0.792	0.642

Table 3. RMSE value for testing data

Clustering Methods	Pneumonia	Pulmonary TB
Subtractive	0.730	0.659
2 MF	0.800	0.529

Representation of the success of the system in the training and testing process is shown in Figure 2 and Figure 3, respectively. The blue circular dot (o) is the true value and the red star dot (*) is the predicted output of the ANFIS model. In the training process, both models provide good results. Most of the model's output values are close to their true values.

In the testing process, some outputs of the both systems have not been able to verge the true value. This is likely because there is an input data whose value is outside the range of the input data training. This unsatisfactory result is also shown by the quite large difference in the value of RMSE training and testing. System performance could be improved by increasing the amount of training.

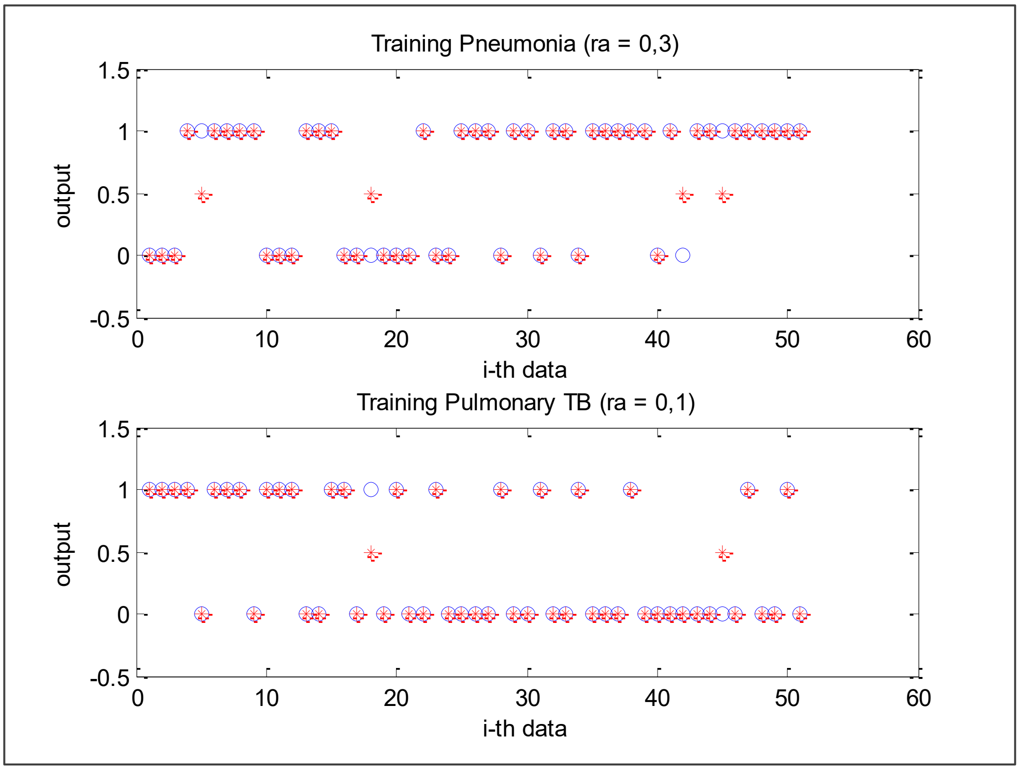

Figure 2. Validation of training process.

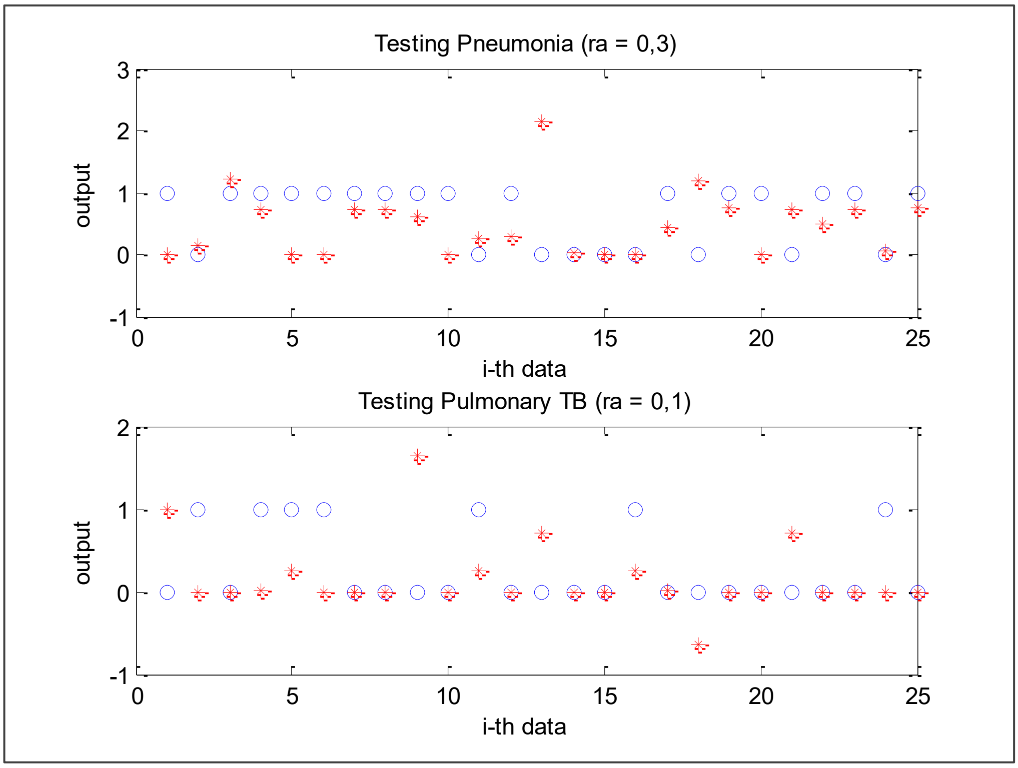

Figure 3. Validation of testing process.

4. Conclusions
The design of the ANFIS model for early detection of pneumonia and pulmonary TB is done by subtractive clustering method. The results of the training showed that both models were able to give better performance compared to the model built using conventional clustering methods. The test results show that both models have comparable performance compared to their counterpart.

5. References
[1] Badan Penelitian dan Pengembangan Kesehatan Kementrian Kesehatan RI 2018 *Hasil Utama Riset Kesehatan Dasar* (Jakarta: Badan Penelitian dan Pengembangan Kesehatan Kementrian Kesehatan RI)
[2] Yunus M and Setyowibowo S 2011 *Jurnal Teknologi dan Informasi* **2** 95–114
[3] Kurnia R, Rahmadewi R and Aini F 2016 *National Conference of Applied Sciences, Engineering, Business and Information Technology* 15–16
[4] Utomo P, Wiharto and Suryani E 2016 *Jurnal Teknologi dan Informasi* **1** 102–106
[5] Asnawati, Bendriyanti R P and Aspriono H 2013 *Jurnal Media Infotama* **9** 162–205
[6] Purwar R K and Srivastava V 2018 *International Journal of Advanced Research in Artificial Intelligence* 99–108
[7] Saputra A 2017 *IT Journal* **4** 109–120
[8] Wanita F 2017 *Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi* **2** 74–79
[9] Rahmadewi R and Kurnia R 2016 *Jurnal Nasional Teknik Elektro* **5** 7–12
[10] Rab T 2010 *Ilmu Penyakit Paru* (Jakarta: Trans info media)
[11] Manurung S 2009 *Gangguan Sistem Pernafasan Akibat Infeksi* (Jakarta: Trans info media)
[12] Jang J S 1993 *IEEE Transactions on Systems, Man, and Cybernetics* **23** 1–21
[13] Chiu S 1996 *Proceedings of North American Fuzzy Information Processing* 461–465