Smart System for Lung Disease Early Detection

N Hasanah¹, F Ariffin², D Irmawati³ and Muslikhin⁴

¹,²,³,⁴ Yogyakarta State University, Karang Malang Campus, Colombo Street No.1, Caturtunggal, Kec. Depok, Kabupaten Sleman, Yogyakarta Special Region 55281, Indonesia.

E-mail: nurhasanah@uny.ac.id

Abstract. Tuberculosis and Pneumonia are two of the ten deadliest diseases in Indonesia, and Chronic Obstructive Pulmonary Disease (COPD) ranks fourth as the leading cause of death in the world. In fact, Indonesian people still lack awareness of COPD. However, some conditions of body organs such as lungs can be diagnosed through sounds. Although pulmonary auscultation has been widely used, technological advancement in the field of digital lung sound analysis and classification provide new possibilities that need to be explored. ANFIS is a popular branch of AI science and is widely applied for classification, prediction, and image processing. ANFIS is essentially able to implement human expertise used in a decision-making process. The objectives of this study are (1) to develop a smart system for the early detection of lung diseases using ANFIS, and (2) to determine smart system accuracy. The smart system was developed using a Waterfall model in sequence starting from analysis, design, coding, and testing. ANFIS that had been built consisted of four inputs and one output, with 81 rules. At the ANFIS development stage, several ANFIS architectures with MF type and different optimization methods had been tested. The smallest RSME value was 4.5357e-06, which was obtained from the training data testing using MF trapezoid (trapmf), hybrid optimization method, error tolerance of 0.0001, and 30 epochs.

1. Introduction

Lung is one of the organs in the respiratory system, which functions as a place for oxygen to exchange with carbon dioxide in blood. Lung diseases have a serious impact on the respiratory system, which can be fatal if not seriously treated. The lung diseases make the sufferers difficult to breathe, to move, and to get oxygen. Even, if not detected quickly, they can cause death [1]. However, there are many people who are less concerned about lung health; this increases the number of lung diseases patients who suffer from Tuberculosis, Bronchitis, COPD (Chronic Obstructive Pulmonary Disease), and Pneumonia [2]. In addition, the lung disease sufferers need to incur high-cost consultation with a specialist [3].

Tuberculosis (TB) becomes a highly contagious disease that leads causes of death in Indonesia. In 2016, there were 274 deaths per day in Indonesia. In the same year, the TB case was suffered by 1,020,000 people. Due to the fact, Indonesia ranks second with the largest number of TB patients in the world after India [4]. Tuberculosis and Pneumonia are the most common causes of death among people in Indonesia [5], while Chronic Obstructive Pulmonary Disease (COPD) ranks fourth as the leading cause of death in the world and is estimated to rank third in 2020. However, many people still lack awareness and social stigma of the disease, so only half of the approximately 210 million people estimated to suffer from COPD have been officially diagnosed. Indonesia is among countries which lack awareness of COPD, because not all sufferers feel or even realize the symptoms [6].
Audio media is a form of a primitive and natural communication that is well known by society. Sounds also can be used to find out expressions of feelings and social interactions. Voice pathology can be associated with respiration, nasal, neural, and laryngeal damages. Based on this, sound medical procedures are important for analyzing and diagnosing a disease. Sound analysis involves a transformation of the sound signals into units and parameters in order to find the characteristics of them. Some conditions of the body organs, particularly lungs, can be diagnosed through sounds. There are not many approaches regarding how to detect normal and abnormal lung sounds. Although pulmonary auscultation has been widely used, the recent technological advancement of hardware, acoustics and computerized lung sound analysis and classification provide new possibilities that need to be further explored.

Neuro-fuzzy is a branch of artificial intelligence that is popular and is widely applied to solve various kinds of everyday problems, especially for classification, prediction, and image processing. This is a combination of the two systems namely fuzzy logic system and artificial neural network. The neuro-fuzzy systems are based on the fuzzy inference systems that are trained using learning algorithms derived from artificial neural network systems. Thus, the neuro-fuzzy system has all the advantages possessed by fuzzy inference systems and artificial neural network systems. The combination of the artificial neural networks and fuzzy logic is essentially capable of implementing human expertise, so it can be used in the decision making process. From its ability to learn, neuro-fuzzy systems are often referred to ANFIS (Adaptive Neuro-Fuzzy Inference System) [7]. ANFIS has been widely used in various fields, including heart disease prediction[8], errors detection and classification in transmission lines [9], robot motor movements (Nickhade et al, 2013), and cash flow forecast [10].

This study aims to develop a smart system for the early detection of lung diseases using ANFIS. The research is expected to find a system that can help detect symptoms of lung diseases with a high level of accuracy.

2. Literature Review

2.1. Lung Diseases
Lung is an organ shaped like large sponges located in the thorax on the other side of the heart and large blood vessels. Lung extends from the root of the neck to diaphragm and is roughly conical with the top in the upper side and the base in the lower side. Heart, large blood vessels, trachea and esophagus, thoracic duct and thymus glands are inside the mediastinum. Lung is divided into lobes. The left lung has two lobes, which are separated by a sloping hemisphere. The superior lobe is located above and in front of the cone-shaped inferior lobe. The right lung has three lobes. The lower lobe is separated by an oblique fissure in the same position as the left inferior lobe. The rest is divided by a horizontal fissure into the upper and middle lobes. Each lobe is further divided into segments called bronco-pulmonary; they are separated by connective tissue walls, each of which is an artery and a vein. Each segment is also divided into units called lobules [11].
The main function of the lungs is to exchange gas in blood and atmosphere. The gas exchange aims to supply oxygen to the tissues and remove carbon dioxide. The need for oxygen and carbon dioxide continues to change according to the level of a person’s metabolism activity, but respiration must keep maintaining the oxygen and carbon dioxide content [13]. The air enters the lungs through a system of narrowed pipes (bronchi and bronchioles) that branch off the two main lungs (trachea). The pipe ends in the lung bubbles (alveoli) as the last air sacs where oxygen and carbon dioxide are transferred from the place where the blood flows. There are more than 300 million elastic alveoli in the human lungs. The air space is maintained in an open condition by surfactant chemicals that can neutralize the tendency of the alveoli to deflate [14].

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)
ANFIS is a combination of the backpropagation neural network and fuzzy logic concepts. Fuzzy based systems can be expressed with the knowledge of "if-then" without requiring mathematical analysis for modeling. Besides, the fuzzy system can also process human reasoning and knowledge, which are oriented on the qualitative aspects [15]. However, artificial neural networks also provide an advantage. It helps classify an object based on a set of features that becomes a system input, because by entering a number of features and doing the training with the data are able to distinguish one object from another [16]. The ANFIS method framework has five layers, namely fuzzification, rule, normalization, defuzzification, and neuro single result [17]. The ANFIS structure is presented in figure 2.

![Figure 1. Lung anatomy [12]](image1)

![Figure 2. ANFIS Structure [18]](image2)
ANFIS is an adaptive neural network based on a fuzzy inference system. An ANFIS hybrid learning procedure can build a mapping of input – output, both of which are based on human knowledge (in the form of fuzzy if–then rules) with the right membership function.

A fuzzy conclusion system that utilizes fuzzy if-then rules can model aspects of qualitative human knowledge and provide reasoning processes without using an appropriate quantitative analysis. Several basic aspects in this approach that require better understanding are presented as follows: 1. There is no standard method for transforming human knowledge or experience into the rule base and database about fuzzy inference systems; 2. There is a need for an effective method for tuning Membership Functions/MF to reduce the size of output errors or to maximize the achievement index.

ANFIS acts as a basis for building a set of fuzzy if-then rules with the right membership function, which is able to produce the right input-output pairs. The fuzzy model can be used as an alternative a multi-layer perceptron. In this case, the system can be divided into 2 groups, namely neural networks with fuzzy weights and fuzzy activation functions, and neural networks with fuzzy inputs in the first and second layer. However, the weights on the neural network are not fuzzy; the neuro fuzzy is in the second group [7].

3. Research Method

Waterfall model is one of the SDLC models, which provides a software life cycle approach sequentially starting from analysis, design, coding, and testing [19]. The illustration of the Waterfall model is shown in Figure 3.

![Waterfall model illustration](image)

The description of the stages in the Waterfall model is as follows: 1. Software needs analysis is a process in identifying needs carried out intensively to specify what the software needs that later will be used by the user; 2. Designing is a multi-step process that focuses on the development of the software programs starting from data structures, software architectures, interface representations, and coding procedures; 3. Coding is a process by which the completed design is then translated into a software program. The result of this coding stage is a computer program matching the design that has been created; 4. Testing is a stage that focuses on the software testing in terms of logic and functional aspects. Also, it ensures that all parts have been tested to minimize errors and the output produced is exactly as expected [19].

The advantages of using a waterfall model are the clear structure of system development stages and the documentation generated at each stage. Also a certain stage will be carried out after the previous stage is completed so there is no overlap in the implementation [19].

4. Findings and Discussion

4.1. Needs Analysis

This stage aims to identify needs that meet the users’ requirements. The observations and interviews were carried out with a pulmonary specialist as the user to find out the symptoms through sounds of lung disease patients.
4.1.1 Analysis of input-output needs. Based on the observation of interviews conducted with a specialist in pulmonary disease, a smart system needs to be made to help doctors detect lung disease in patients. This smart system can function as a doctor's assistant to detect the lung disease early. This smart system can also be used by the public. This smart system can detect early lung disease with human breath sounds as the input. The output is a statement (normal or abnormal condition) which is based on the lung disease detection.

4.1.2 Process needs analysis. This stage consists of two processes: (1) collecting the sound samples for the ANFIS knowledge base, and 2) determining a certain type of lung the diseases that have been inputted. In the first process, the researchers took data in the form of breath sounds of lung disease patients. The types of pulmonary diseases identified based on the breath sound samples were Bronchiolitis, Bronchiectasis and Pulmonary Edema. The normal patient's breath is also included. As many as 48 data of the pulmonary disease patients' breath sounds were taken in the “.wav” file format. The breath sound data would then be included in ANFIS as a knowledge base to become the basis to make a decision. Furthermore, the data would be divided into two, namely training data and testing data.

4.2. Design
The next stage was to design the application interface of the smart system for the early detection of lung diseases.

4.3. Implementation
The application interface is shown in Figure 4. There are “open file” and “play” buttons. After the user has selected a sound file, the user can immediately see the graph. If the user wants to listen to his voice, the user can choose the “play” menu. In addition, there is also a “predict” menu which functions to detect the lung disease based on sound files that have been inputted by the user. The result of the sound identification from the “broncl.wav” file that the user chose can be seen in Figure 4. The lung disease that has been identified is bronchiolitis.

Figure 4. Application Interface of smart system for early detection of lung diseases
4.4. Testing
At this stage, the information was collected from users who had used the application. The experts appointed had mastered the software development and those who were involved in the testing were potential users.

4.5. Needs Analysis
This study employed two types of data, namely training data (35) and testing data (13).

4.5.1 Data Loading. At this stage, the data was loaded into ANFIS Editor GUI. The training data was loaded and then followed by testing data. The result of the training data loaded is presented in Figure 5, which is depicted in small circles with holes.

Moreover, the testing data result that has been loaded is presented in Figure 6 which is depicted with small dots filled. Both types of data loaded are in the form of sound files with the extension of “.dat”.

![Figure 5. Training data](image-url)
4.5.2 FIS generation. The next step was to generate FIS. In the input, the number of Membership Function (MF) used was 3 3 3 3, while the MF types that would be used were triangle (trimf) and trapezoid (trapmf). At the output, the MF type selected was constant.

4.5.3 FIS Training. After generated, the FIS was trained as the next step. The optimization methods that would be used were hybrid and backpropagation. The error tolerance was 0.0001 and the epochs used were 30.

Figure 6. Testing data

Figure 7. Training FIS

Figure 7 shows that the FIS training phase uses hybrid and trapmf methods. The Root Mean Square Error (RMSE) value obtained is 4.6798e-06.
4.5.4 FIS Testing. At this stage, testing was carried out using the FIS design that had been made previously on the training data continued by the testing of the other data. This test was done using two types of MF: trimf and trapmf. Whereas the optimization methods used were hybrid and backpropagation. Figure 8 shows the results of the training data testing using trapmf and hybrid optimization methods. The RMSE value obtained is 4.5357e-06.

![Figure 8. Testing result of training data](image1)

Based on Figure 9, the RMSE value obtained from the testing data is 0.060176.

![Figure 9. Testing result of testing data](image2)

Table 1 shows the comparison of RSME values obtained from the testing results of the training and testing data, using two types of MF (trimf and trapmf) and optimization methods (hybrid and backpropagation).
Table 1. Results of RSME Value Comparison

| Membership Function (MF) | Training Data | Testing Data |
|--------------------------|---------------|--------------|
|                          | Hybrid        | Backpropagation | Hybrid | Backpropagation |
| Trimf                    | 0.00020592    | 0.3148        | 0.037486 | 0.35992 |
| Trapmf                   | **4.5357e-06** | 0.30487       | 0.060176 | 0.3249 |

Based on the Table 1, it is concluded that the smallest RMSE result is 4.5357e-06, which is obtained from the testing results of the training data, using trapmf with the hybrid optimization method. In addition, figure 10 shows the structure of the ANFIS model used. The ANFIS consists of four inputs and one output, with 81 rules.

Figure 10. ANFIS model structure

The following is a summary of the ANFIS architecture used to obtain the smallest RSME.
- **MF Type**: trapmf
- **The number of MF**: 3 3 3 3
- **Optimization method**: hybrid
- **Error Tolerance**: 0.0001
- **Epochs**: 30
- **Rules**: 81

5. Conclusions and Suggestions

5.1. Conclusions
The smart system for the early detection of lung diseases is developed using the Waterfall model. ANFIS that has been built for this application consists of four inputs and one output, with 81 rules. At the ANFIS development stage, several ANFIS architectures have been tested using MF type and different optimization methods. The lowest RSME value is 4.5357e-06, obtained from the “training data” using MF trapezoid (trapmf) and hybrid optimization methods, error tolerance of 0.0001, and 30 epochs.
5.2. Suggestions
Suggestions for further research: 1. The MF type is added so that it can be compared to obtain the minimum RMSE value; 2. The input variables in FIS can be added so that the detection results of lung diseases can be more accurate; 3. The portable and real time application can be developed further.

6. References

[1] A Saputra 2011 Sistem Pakar Identifikasi Penyakit Paru-Paru Pada Manusia Menggunakan Pemrograman Visual Basic 6.0 Jurnal Teknologi Dan Informatika (Teknomatika) Vol 1 No 3
[2] I Junaidi 2010 Penyakit Paru dan Saluran Napas (Jakarta: Bhuana Ilmu Popular)
[3] I T Dessetiadi, A Pujianto, and M G Ardi 2016 Sistem Pakar untuk Mendiangosa Penyakit Paru-Paru Menggunakan Algoritma Bayes Seminar Nasional Teknologi Informasi dan Multimedia 2016
[4] L M Putra 2018 Indonesia Peringkat Kedua TBC di Dunia, Waspadai Gejalanya accessed from https://sains.kompas.com/read/2018/01/17/070700923/indonesia-peringkat-kedua-tbc-di-dunia-waspadai-gejalanya
[5] U Widowati 2015 10 Penyakit Paling Mematikan di Indonesia accessed from https://www.cnnindonesia.com/gaya-hidup/20150513163407-255-53129/10-penyakit-paling-mematikan-di-indonesia
[6] Tempo.co 2017 Kesadaran terhadap Penyakit Paru Obstruktif Kronis Masih Rendah accessed from https://www.msn.com/id-id/kesehatan/health-news/kesadaran-terhadap-penyakit-paru-obstruktif-kronis-masih-rendah/ar-BBFbjRC?li=AA4WXs
[7] S Kusumadewi and S Hartati 2010 Neuro-Fuzzy Integrasi Sistem Fuzzy dan Jaringan Syaraf (Yogyakarta: Graha Ilmu)
[8] N Ziasabounchi and I Askernzade 2014 ANFIS Based Classification Model for Heart Disease Prediction International Journal of Electrical & Computer Sciences IJECS-IJENS Vol 14 p 7-12
[9] Isa A, Verma H, Danladi H and Suleiman B 2016 NFIS-Based Algorithms for Detection and Classification of Fault on Transmission Lines European Journal of Advances in Engineering and Technology Vol 3 No 8 p 47-53
[10] Wang J and Ning C 2015 ANFIS Based Time Series Prediction Method of Bank Cash Flow Optimized by Adaptive Population Activity PSO Algorithm Information Vol 6 p 300-313
[11] Snell R 2006 Anatomi Klinik Untuk Mahasiswa Kedokteran Ed. 6 (Jakarta: EGC)
[12] Watson R 2002 Anatomi Dan Fisiologi Edisi 10 (Jakarta: Buku Kedokteran ECG) p 303
[13] West J W 2004 Interaction of Energy and Bovine Somatotropin with Heat Stress Journal of Dairy Science Vol 43 p 1245
[14] McArdle WD 2006 Exercise Physiology: Energy, Nutrition, and Human Performance. 4 th Edition, (USA: Williams and Wilkins) p 19-41
[15] Jang J S R 1993 ANFIS: Adaptive-Network-Based Fuzzy Inference Systems IEEE Trans. on Systems, Man and Cybernetics Vol 23 No 3 p 665-685
[16] Duda R O, Hart P E, and Stork D G 2001 Pattern Classification (New York: John Wiley & Sons)
[17] Ata R and Kocyigit Y 2010 An Adaptive Neuro Fuzzy Inference System Approach for Prediction of Tip Speed Ratio in Wind Turbines Expert System with Applications Vol 37 p 5454-5460
[18] Wu J D, Hsu C C, and Chen H C 2009 An Expert System of Price Forecasting for Used Cars Using Adaptive Neuro Fuzzy Inference Expert System with Applications Vol 36 p 7809-7817
[19] A S Rosa and Shalahuadin 2013 Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek (Bandung: Informatika)