An application of ANFIS for lung diseases early detection system

Mochamad Yusuf Santoso*[1], Am Maisarah Disrinama[2], Haidar Natsir Amrullah[3]
Occupational Health and Safety Engineering, Shipbuilding Institute of Polytechnic Surabaya, Indonesia[1,2,3]

Abstract
Indonesian Basic Health Research in 2018 showed the prevalence of pneumonia, pulmonary tuberculosis (TB) and lung cancer in Indonesia 4.0% 0.4% and 0.18%, respectively. However, the number of lung specialists is small. According to the Indonesian Lung Specialist Association webpage, the number of doctors joined in the association up to 2008 were 452. This amount is very less when compared with existing lung disease cases. Thus, the handling of lung disease will be too late. The use of ANFIS for early detection of lung disease is growing. However, the systems designed are need preprocessing data to be executed and still applied for one type of disease. This paper will design a desktop application based on ANFIS expert system to detect lung disease early, i.e. for pneumonia, pulmonary TB and lung cancer. The system will work based on simple symptoms expressed by the patient. Subtractive clustering is used for clustering process. The results of the training showed that the models were able to give better performance compared to the model built using conventional clustering methods. The test results show that those three models have comparable performance compared to their counterpart. Software validation shows that the it gives 94.00% succeed for training data and up to 100% for testing data. This application is not intended to replace the role of a doctor, but to help diagnose the patient's condition earlier.

1. Introduction

Basic Health Research in 2018 showed the prevalence of pneumonia, pulmonary tuberculosis (TB) and lung cancer in Indonesia is 4.0%, 0.4% and 0.18%, respectively [1]. Those disease are three of the ten most common diseases in Indonesia [2]. Serious attention is needed because lung disease in Indonesia is quite high. 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 [3]. To date, pulmonary disorders detections are generally carried out by examination of physical symptoms by doctors [4]. The lung disease sufferers need to incur high-cost consultation with a specialist [5]. However, with a minimal number of lung specialists, the handling of lung disease will be too late. According to the Indonesian Lung Specialist Association (PDPI) webpage, the number of doctors joined in the association up to 2008 was 452 doctors [6]. This amount is very less when compared with existing lung disease cases [7].

The expert system existence is currently needed [8]. It helps the community in the absence of an expert [9]. The expert system in health is not intended to replace the role of a doctor, but to help diagnose the patient's condition earlier [10]. Artificial Intelligence Techniques for Cancer Detection and Classification have been investigated [11]. 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 [12]. Simulations of early detection of pulmonary disease based on patient symptoms by the Bayesian method [13] and X-ray image analysis based on image processing were successfully carried out [4]. Combination of image processing and artificial neural network can be utilized for breast cancer early detection [14]. Breath phase detection based on acoustic and non-acoustic approaches are provided in [15].

Expert systems in health have been growing lately, one of which is the Adaptive Neuro Fuzzy Inference System (ANFIS) [16]. High detection accuracy for breast cancer early detection was obtained using ANFIS [17]. Lung cancer classification using ANFIS gives better performance compared with Fuzzy Inference System (FIS) [18]. The use of ANFIS for lung disease detection is growing. However, the systems need preprocessing data for ANFIS modelling. Image processing of computer tomography (CT) images of chest were applied for lung cancer staging [19], lung cancer recognition [20], pulmonary TB diagnosis [21] and Chronic Obstructive Pulmonary Disease (COPD) screening [22]. Besides, those designed systems are still put in an application for one type of disease.

This paper will design an application based on ANFIS to detect lung diseases early, i.e. for pneumonia, pulmonary TB and lung cancer, using simple symptoms from patient. The modelled ANFIS expert system will be transformed to user friendly desktop application. Hopefully, after 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 or even none. With this system, it is expected that more and more patients will be detected earlier about the illness, so that the treatment will be made easier.

The rest of the paper is structured as follows. Section 2 describes proposed system design, including methods for testing and validating the proposed system. Section 3 gives the results and explanations of test and validation while Section 4 concludes the paper.

2. Research Method
2.1 Lung Disease Symptoms

Lung disease can be classified as infectious and non-communicable. Infectious pulmonary diseases include pneumonia and pulmonary TB and lung cancer as non-communicable [1]. Pneumonia and pulmonary TB have some of the same symptoms. However, there are some more specific symptoms that make handling differently for each disease [23]. The main symptom for lung cancer is cough, but it also can be detected by patient behavior [24]. Pneumonia, TB and lung cancer symptoms that are used as input for the ANFIS model to be designed are shown in Table 1.

Table 1. Lung Disease Symptoms

Lung Disease	Symptoms
Pneumonia	Cough Fever Headache Malaise Shivering
Pulmonary TB	Cough Fever Malaise Diaphoresis Shivering
Lung Cancer	Cough length Type of work Smoking duration Number of cigarettes

2.2 Adaptive Neuro Fuzzy Inference System

Adaptive Neuro Fuzzy Inference System (ANFIS) is a fuzzy inference system (FIS) which implemented in the framework of neural network [25]. It uses the first order Takagi-Sugeno-Kang (TSK) model, for computation simplicity and convenience [26]. 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.
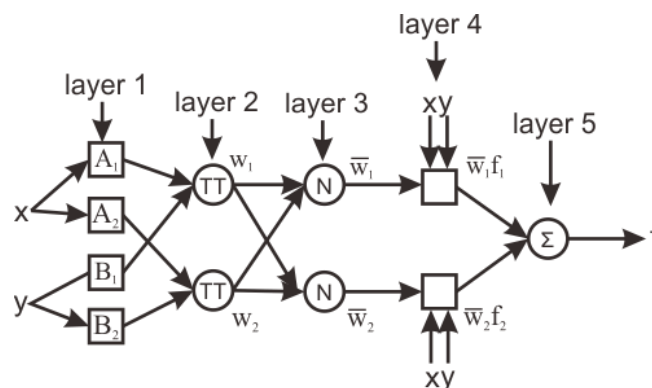

Figure 1. ANFIS Structure [26]

Mathematical equation for Layer 1 depends on type of membership function (MF). Equation 1 and Equation 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.

Every node in Layer 3 is non adaptive node that show normalized firing strength of the node, given in Equation 4. In Layer 4, represented by Equation 5, 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 with function of Equation 6. The output is a decision from the designed system.

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

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

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

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

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

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

This paper designs three ANFIS structures for pneumonia, pulmonary TB and lung cancer 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. ANFIS system for lung cancer has four inputs, i.e. cough duration, type of work, smoking duration and the number of consumed cigarettes. 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. For lung cancer, there are 75 patient data: 50 for the training and the rest are for testing process. The data are collected from questionnaire. Each patient with cough as the main keluhan answered the questions. The questionnaire consists of linguistic and numbering answers. The linguistic answers then converted to number for ANFIS modelling. Table 2 gives the data characteristics for each question. The doctor then validated the questionnaire by giving the output, i.e. risk or not.

2.3 Subtractive Clustering

Subtractive clustering method was proposed in [27]. The method makes each data points are considered as the candidates for cluster center. In subtractive clustering, a data point with the highest potential, which is a function of the distance measure, is considered as a cluster center. 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 $\alpha = \frac{\gamma}{r_a^2}$

P_i 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 neighborhood range of the cluster or simply the radius of hypersphere cluster in data space.

Subtractive clustering has four parameters, namely, accept ratio $\bar{\varepsilon}$, 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 $\bar{\varepsilon}$ and *ε* will result in small number of rules. Conversely, small values of $\bar{\varepsilon}$ 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 r_a 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 the smallest Root Mean Square Error (RMSE).

Table 2. Data Characteristics and its Conversion

Question	Answer options	Conversion
Cough	no	0
	phlegm is hard to get out	1
	phlegm comes out easily	2
	thick phlegm	3
	dry cough	4
	bleeding cough	5
	coughing heavier	6
	reduced cough history	7
Cough length	number	number
Fever	no	0
	goes up and down	1
	prolonged fever	2
	fever at night	3
	convulsions	4
Fever length	number	number
Headache	no	0
	spinning	1
	forehead dizzy throbbing pain	2
	stiff neck	3
	moving around	4
	as if drilled	5
Malaise	no	0
	lazy to do activities	1
	lost work mood	2
Diaphoresis	no	0
	all day long	1
	at night	2
Shivering	no	0
	all day long	1
	at night	2
	when have a fever	3
Type of work	Housewives, Teachers/Lecturers, Office Employees, Bank Tellers, Home Food Sellers, Estate Officers, etc.	1
	Traveling Traders, Woodworkers, Drivers, Ceramic Craftsmen, etc.	2
	Mining Workers, Chemical Plant Employees, Other Field Workers	3
Smoking duration (years)	number	number
Number of cigarettes	number	number

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 r_a value of 0.1. Lung cancer ANFIS model has best RMSE with cluster radius value of 0.4. Table 3 shows the results of the RMSE calculation for each model. Each is then compared with a model 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 [27].

Table 3. RMSE Value for Training Data

Clustering Methods	Pneumonia	Pulmonary TB	Lung Cancer
Subtractive	0.140	0.099	1.42E-15
2 MF	0.792	0.642	0.721

The testing data is used to test the selected model. The RMSE results for the lung disease early detection system are shown in Table 4. The early detection system for pneumonia and lung cancer using subtractive give better performance compared to models built using conventional clustering methods. For early pulmonary TB detection systems, both models have comparable performance.

Table 4. RMSE Value for Testing Data

Clustering Methods	Pneumonia	Pulmonary TB	Lung Cancer
Subtractive	0.730	0.659	0.348
2 MF	0.800	0.529	0.600

Figure 2 and Figure 3 represent the success of the training and testing process for the pneumonia and pulmonary TB system, 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. Regarding Figure 4 for lung cancer ANFIS model, it has promising training validation result.

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 can be improved by increasing the amount of training data [28].
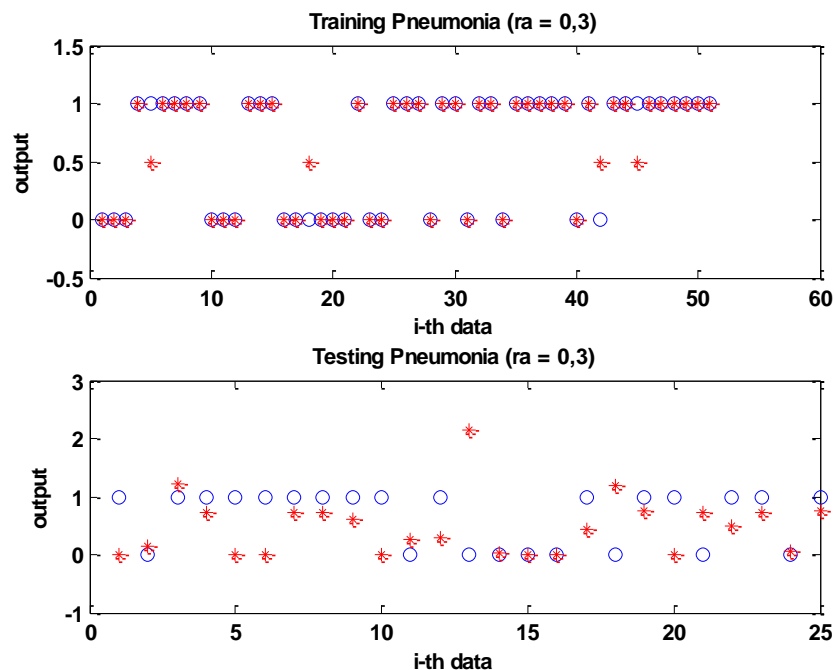

Figure 2. Representation of Training and Testing Process for Pneumonia ANFIS Model

The selected ANFIS models are converted into a lung disease early detection software. The success rates for each disease detection are given in Table 5. The software has satisfactory results. Using training data, the software has performance up to 94% succeed. Higher success rates are obtained as testing data are entered to the software. Overall, early detection for lung cancer has the best performance.
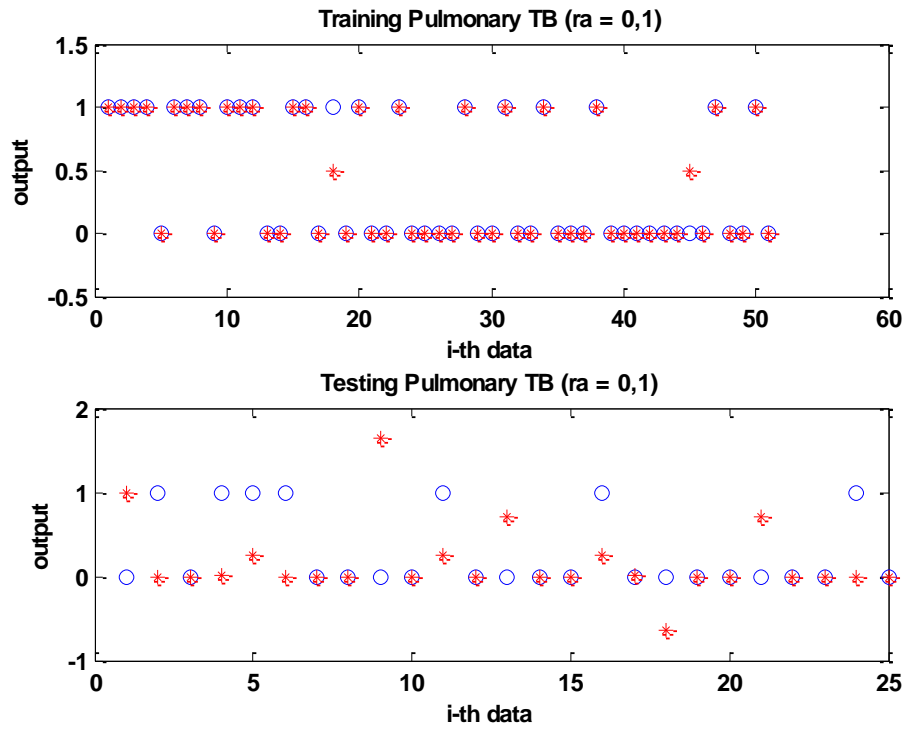
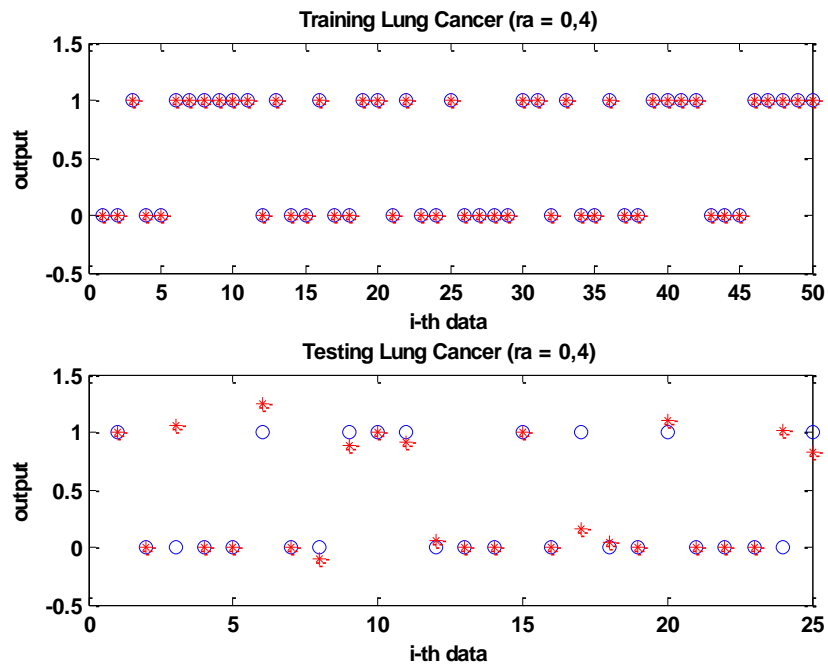Figure 3. Representation of Training and Testing Process for Pulmonary TB ANFIS Model

Figure 4. Representation of Training and Testing Process for Lung Cancer ANFIS Model

Table 5. Validation Results for Lung Disease Early Detection Software

Lung Disease	Success Rate (%)	
	Training Data	Testing Data
Pneumonia	86.27	96.00
Pulmonary TB	72.54	80.00
Lung Cancer	94.00	100.00

4. Conclusion

The ANFIS models for early detection of lung disease was designed by subtractive clustering method. The training process results showed that those three models were able to give better performance compared to the model built using conventional clustering methods. The results of testing process show that those three models have comparable performance compared to their counterpart. The desktop application validation shows that the it gives satisfactory results with up to 94.00% succeed for training data and up to 100% for testing data.

Notation

μ : membership function value.
c : gaussian function mean
σ : gaussian function deviation
w : weighting value
$\bar{w}$: averaged weighting value

References

[1] Badan Penelitian dan Pengembangan Kesehatan Kementrian Kesehatan RI, "Hasil Utama Riset Kesehatan Dasar," 2018.
[2] D. Purnamasari, "The Emergence of Non-communicable Disease in Indonesia," Acta Med. Indones., Vol. 50, No. 4, Pp. 273–274, 2018.
[3] M. Yunus and S. Setyowibowo, "Aplikasi sistem pendukung keputusan diagnosa penyakit paru- paru dengan metode forward chaining," J. Teknol. Inf. Teor. Konsep, dan Implementasi, Vol. 2, No. 2, Pp. 95–114, 2011.
[4] R. Rahmadewi and R. Kurnia, "Klasifikasi penyakit paru berdasarkan citra rontgen dengan metoda segmentasi sobel," J. Nas. Tek. Elektro, Vol. 5, No. 1, Pp. 7–12, 2016. https://doi.org/10.20449/jnte.v5i1.174
[5] I. T. Dessetiadi, A. Pujianto, and M. G. Ardi, "Sistem Pakar Untuk Mendiagnosa Penyakit Paru-Paru," in Seminar Nasional Teknologi Informasi dan Multimedia 2016, 2016, Pp. 3.4-25-3.4-30.
[6] http://www.klikpdpi.com/index.php?mod=content&sel=18
[7] P. Utomo, Wiharto, and E. Suryani, "Sistem Diagnosa Penyakit Paru Berdasarkan Foto Rontgen Dengan Pendekatan Fuzzy Learning Vector Quantization," J. Teknol. Inf. ITSmart, Vol. 1, No. 2, Pp. 102–106, 2016. https://doi.org/10.20961/its.v1i2.604
[8] R. Fauzan and A. V. Prananda, "Expert System for Diagnosing Palm Tree Diseases and Pests using Forward Chaining and Certainty Factor," Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control, Vol. 3, No. 1, Pp. 27–34, 2018. https://doi.org/10.22219/kinetik.v3i1.524
[9] Asnawati, R. P. Bendriyanti, and H. Aspriono, "Sistem Pakar Mendeteksi Penyakit Asma Pada Puskesmas Lingkar Timur Bengkulu," J. Media Infotama, Vol. 9, No. 2, Pp. 162–205, 2013.
[10] A. Saputra, "Pengembangan Sistem Pakar Identifikasi Penyakit Paru-Paru Menggunakan Metode Certainty Factor," IT J., Vol. 4, No. 2, Pp. 109–120, 2017.
[11] A. R. M. Al-shamasneh and U. H. B. Obaidellah, "Artificial Intelligence Techniques for Cancer Detection and Classification : Review Study," Eur. Sci. J., Vol. 13, No. 3, Pp. 342–370, 2017. https://doi.org/10.19044/esj.2017.v13n3p342
[12] F. Wanita, "Sistem pakar deteksi dini penyakit dengan gejala sesak nafas menggunakan metode forward chaining," in Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi, 2017, Vol. 2, No. 2, Pp. 74–79.
[13] R. Kurnia, R. Rahmadewi, and F. Aini, "Deteksi Dini Penyakit Paru Dengan Metoda Bayesian," in National Conference of Applied Sciences, Engineering, Business and Information Technology, 2016, Pp. 15–16.
[14] M. M. Mehdy, P. Y. Ng, E. F. Shair, N. I. Saleh, and C. Gomes, "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer," Comput. Math. Methods Med., Vol. 2017, 2017. https://doi.org/10.1155/2017/2610628
[15] R. Palaniappan, K. Sundaraj, and F. G. Nabi, "An Overview of Breath Phase Detection – Techniques & Applications," J. Telecommun. Electron. Comput. Eng., Vol. 10, No. 2, Pp. 33–36, 2015.
[16] R. K. Purwar and V. Srivastava, "Recent Advancements in Detection of Cancer Using Various Soft Computing Techniques for MR Images," in Progress in Advanced Computing and Intelligent Engineering, 2018, Pp. 99–108. https://doi.org/10.1007/978-981-10-6872-0_10
[17] P. Zarbakhsh, A. Abdoljalil, and D. Hasan, "Early detection of breast cancer using optimized ANFIS and features selection," in 9th International Conference on Computational Intelligence and Communication Networks, 2017, Pp. 39–42. https://doi.org/10.1109/cicn.2017.8319352
[18] K. Jinsa and G. K., "Lung cancer classification using fuzzy logic for CT images," Int. J. Med. Eng. Informatics, Vol. 7, No. 3, Pp. 233–249, 2015. https://doi.org/10.1504/ijmei.2015.070128
[19] M. Lavanya and P. M. Kannan, "Lung cancer segmentation and diagnosis of lung cancer staging using MEM (modified expectation maximization) algorithm and artificial neural network fuzzy inference system (ANFIS)," Biomed. Res. An Int. J. Med. Sci., Vol. 29, No. 14, 2018. https://doi.org/10.4066/biomedicalresearch.29-18-740
[20] M. Obayya and M. Ghandour, "Lung Cancer Recognition using Radon Transform and Adaptive Neuro Fuzzy Inference System," Int. J. Comput. Appl., Vol. 124, No. 2, Pp. 25–30, 2015. https://doi.org/10.5120/ijca2015905373
[21] B. Ashadevi, P. M. Selvi, and B. S. Revathi, "An Effective Diagnosis of Pulmonary Tuberculosis using K-Means Clustering and ANFIS," Sch. J. Eng. Technol., Vol. 5, No. 8, Pp. 427–439, 2017.
[22] S. K. Meenakshi and R. C.S., "An Efficient ANFIS based Approach for Screening of Chronic Obstructive Pulmonary Disease (COPD) from Chest CT Scans with Adaptive Median Filtering," Asian J. Sci. Res., Vol. 7, No. 1, Pp. 18–32, 2014. https://doi.org/10.3923/ajsr.2014.18.32
[23] S. Smilji et al., "Differences and similarities between the symptoms and clinical signs in patients with pulmonary tuberculosis and pneumonia," Mil. Med. Pharm. J. Serbia, Vol. 76, No. 2, Pp. 192–201, 2019. https://doi.org/10.2298/vsp170301080s
[24] C. E. Mosher, M. A. Ott, N. Hanna, S. I. Jalal, and V. L. Champion, "Coping with Physical and Psychological Symptoms: A Qualitative Study of Advanced Lung Cancer Patients and their Family Caregivers," Support Care Cancer, Vol. 23, No. 7, Pp. 2053–2060, 2016. https://doi.org/10.1007/s00520-014-2566-8

[25]   N. Walia, H. Sing, and A. Sharma, "ANFIS : Adaptive Neuro-Fuzzy Inference System- A Survey," *Int. J. Comput. Appl.*, Vol. 123, No. 13, Pp. 32–38, 2015. https://doi.org/10.5120/ijca2015905635

[26]   J. S. Jang, "ANFIS : Adaptive Network-Based Fuzzy Inference System," *IEEE Trans. Syst. Man. Cybern.*, vol. 23, No. 3, Pp. 1–21, 1993. https://doi.org/10.1109/21.256541

[27]   S. Chiu, "Method and software for extracting fuzzy classification rules by subtractive clustering," in *Proceedings of North American Fuzzy Information Processing*, 1996, Pp. 461–465. https://doi.org/10.1109/nafips.1996.534778

[28]   S. Dalecky and F. V. Zboril, "An Approach to ANFIS Performance," in *Advances in Intelligent Systems and Computing*, 378th ed., Springer, Cham, 2015. https://doi.org/10.1007/978-3-319-19824-8_16