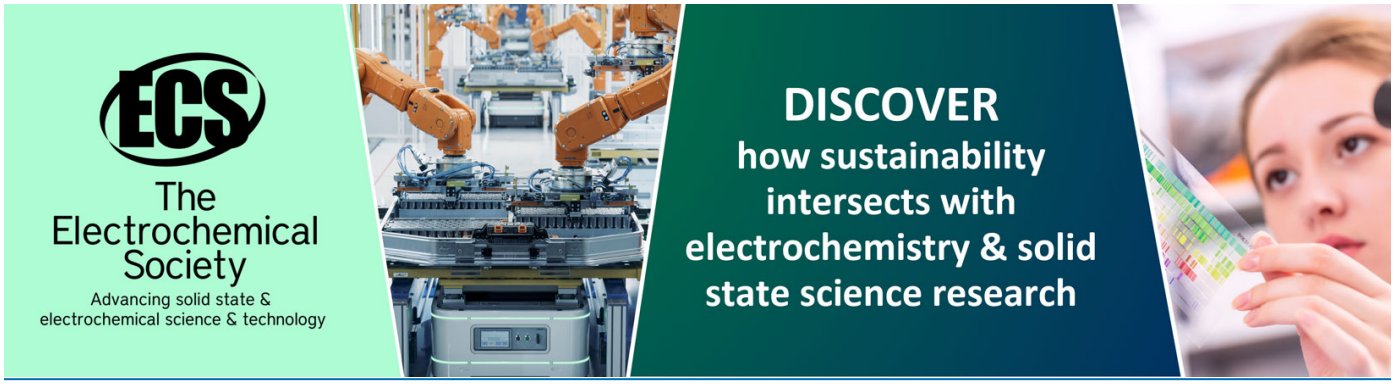
Identification of Tuberculosis (TB) Disease Based on Lung X-Rays using Extreme Learning Machine



Identification of Tuberculosis (TB) Disease Based on Lung X-Rays using Extreme Learning Machine

M K M Nasution[1], M Elveny[1*], S M Hardi[1], I Jaya[1] and M A Siregar[1]

[1] Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Padang Bulan 20115 USU, Medan, Indonesia.

*Email: marischaelveny@usu.ac.id

Abstract *Tuberculosis* is a disease caused by a bacterium called *Mycobacterium Tuberculosis*. Most Tuberculosis bacteria not only attack the lungs, but also attack other organs. *Tuberculosis* disease can spread and contaminate through the salivary and bacterial when the sufferer coughs and sneezes. Thus, the contagion of tuberculosis can take place either through contact directly with the sufferer or through the air. The examination system of Tuberculosis is still done manually. Therefore, a computational method is necessary to examine and identify Tuberculosis suffered by the patient through X-ray images of human lungs. The method used in this research is the Extreme Learning Machine. Lung X-ray images are used as input for image processing. The steps taken before identification were image pre- processing and *Thresholding*. The external results of this research include HKI in the form of software (registered), international conference presenter (done). In addition, the results of this study can assist public in determining the types of tuberculosis suffered so that it will be more efficient in early treatment of the disease.

1. Introduction

Every year in Indonesia there are 583 new cases of tuberculosis with 130 deaths due to positive Tuberculosis. And the number of deaths of 105,952 people is estimated to occur every year in Indonesia. This is also caused by weak socio-economic conditions, poor immune system or poor nutrition and cleanliness of the living environment so that tuberculosis cases occur a lot.

The number of tuberculosis patients in Indonesia is numbered the third highest after India and China with the number of patients about 10% of the total number of tuberculosis patients in the world. Based on National Household Health Survey in 2001 it was found that diseases of the respiratory system were the second leading cause of death after the circulation system. In 1992 National Household Health Survey data mentioned tuberculosis being the second leading cause of death while in National Household Health Survey data in 2001 mentioned that tuberculosis was the first cause of death in the infectious disease class. Nationally, in 2010 pulmonary tuberculosis has attacked the community with 302,861 people. And about 75% of patients suffering from tuberculosis are a group of people who are productive age that is 15-50 years [2]. The results of the data found in Banten province in 2010 were 13,877 cases of pulmonary tuberculosis patients or around 75.2% while in 2013 cases of pulmonary tuberculosis were recorded as many as 5,123 patients in Serang City [2].

TB prevention efforts have been carried out through various health programs at Community Health Center, in the form of developing a TB prevention strategy known as the DOTS strategy (directly observed treatment, short course). MDR (multi-drug resistance) -TB, but the results are still not felt as expected. Therefore, it is hoped that there will be attention from related parties in the effort to increase the involvement of TB care services in the future. Transmission and eradication of pulmonary TB is also inseparable from the socio-cultural aspects of the community concerned. Besides that, health workers such as doctors are expected to always increase their knowledge and skills to be more perfect to detect and diagnose TB at an early stage. [13].

Based on the background above, the researcher will submit research that can help and simplify the process of identifying tuberculosis based on the processing of human X-ray images using the Extreme Learning Machine method.

2. Literature Review

Tuberculosis is a disease caused by a bacterium called Mycobacterium tuberculosis, which is an aerobic bacterium that can live in the human body, especially in the lungs or in other organs that have high oxygen partial pressure. Pulmonary tuberculosis is a serious disease, especially for infants or young children, children with malnutrition, and children with immunological disorders.

The simple symptoms of tuberculosis are chronic cough with phlegm mixed with blood, fever, night sweats and weight loss. Through X-ray photographs, the lungs which are affected or have contracted tuberculosis will have white patches on several parts of the lung, fine infiltrates with a size of several mm spread over the top of both lungs (radiopaedia.org) in contrast to normal lung conditions.

3. Material and Method

This study proposes a system design using image processing that starts from the selection stage which is divided into two datasets, namely: training data and testing data. Training data is data that has known labels and then the data is used as a comparison of testing data in the process of classifying tuberculosis, while testing data is data to find out whether or not tuberculosis into this form is an image processing stage that aims to produce a better image for subsequent stages of processing. The preprocessing stage consists of the vanes and grayscale. Grayscale stage aims for the uniform gray color in the image to be processed. In the original image looks uneven gray color. The next stage uses scaling. This stage is needed to adjust the pixel size in the image, the more number of pixels the more time for image processing.

The stage after preprocessing is *thresholding* which aims to produce an image in black and white. After *thresholding*, the image will be converted into a binary image. The output produced at this stage is 0 and 1. This binary image is used as input in the classification process using extreme learning machines.

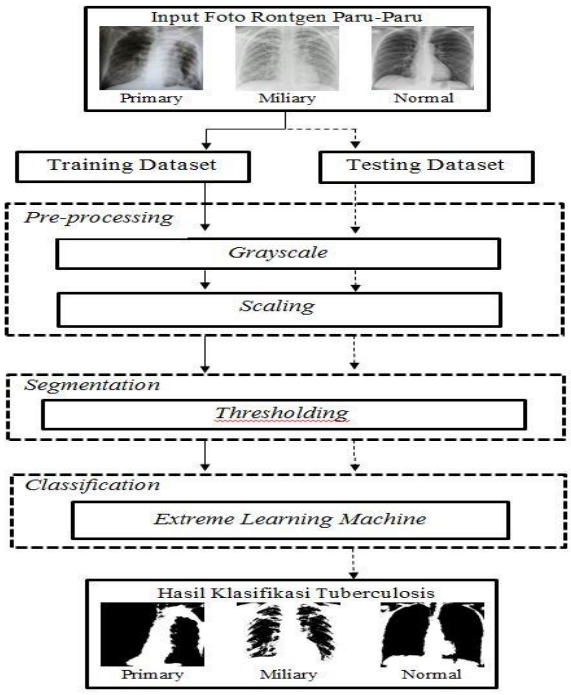


Figure 1. Research Design

Training data is data which has known the labels and is used as a comparison of testing data in the process of classifying tuberculosis. While testing data is data to determine the classification of tuberculosis. Distribution of training data and testing data can be seen in Table 3.1.

Table 3.1. Distribution of Training Data and Testing Data

No	Dataset	The Amount of Data
1	Training Data	45
2	Testing Data	15

The types of tuberculosis used in this study are primary, miliary and normal. The division of images is shown in Table 3.2.

Table 3.2. Distribution of Image based on Disease

No	Dataset	Jumlah Data
1	*Primary*	20
2	*Miliary*	20
3	*Normal*	20

The first stage is grayscale. This stage aims to uniform gray color in the image to be processed. In the original image looks uneven gray color. The formula for calculating the composition of color values is used by reporting the threshold value as in equation 3.1.

Grayscale = (0,3*R + 0,5*G + 0,2B).....(3.1)

Annotations:

R = Red element

G = Green element

B = Blue Element

Image of *grayscale* can be seen in figure 3.2.

Figure 3.2 Image of *Grayscal*e

The next stage uses *scaling.* This stage is required to adjust the pixel size in the image. The more number of pixels, the more time for image processing. Scaling image can be seen in Figure 3.3.

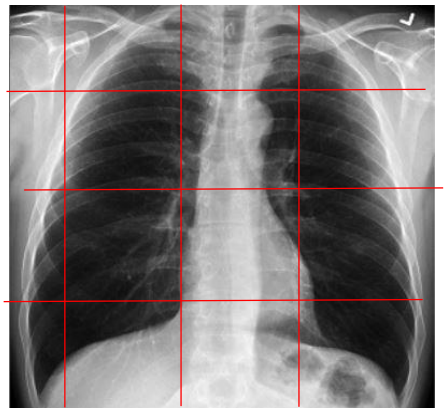


Figure 3.3 Image of *Scaling*

For example, the image which was originally sized 300x300 changed to 4x4. In this study, the image size of 300x300 was changed to 30x30. The author has tested and selected 30x30 sizes according to system requirements.

The stages after preprocessing are thresholding which aims to produce an image in black and white. The results of the thresholding process can be seen in Figure 3.4.

Figure 3.4 Image of *Thresholding*

After *thresholding*, the image will be converted into a binary image. The output produced at this stage is 0 and 1. This binary image is used as input in the classification process using extreme learning machines. Binary image can be seen in Figure 3.5.

Figure 3.5 Image of Binary 30x30

In this study, the method used for the classification process is extreme learning machine. The extreme learning machine architecture for tuberculosis classification can be seen in Figure 3.6.

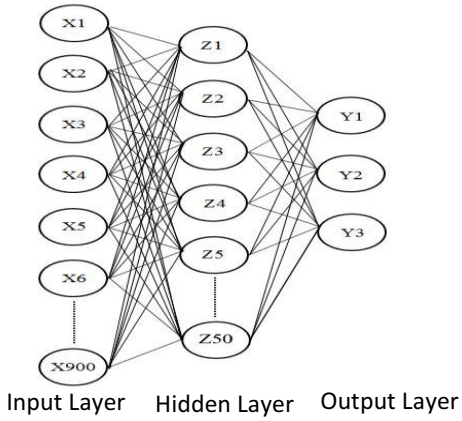


Figure 3.6 Architecture of *Extreme Learning Machine* for classifying *Tuberculosis*

Information from the extreme learning machine architecture images for the classification of tuberculosis can be seen in Table 3.7.

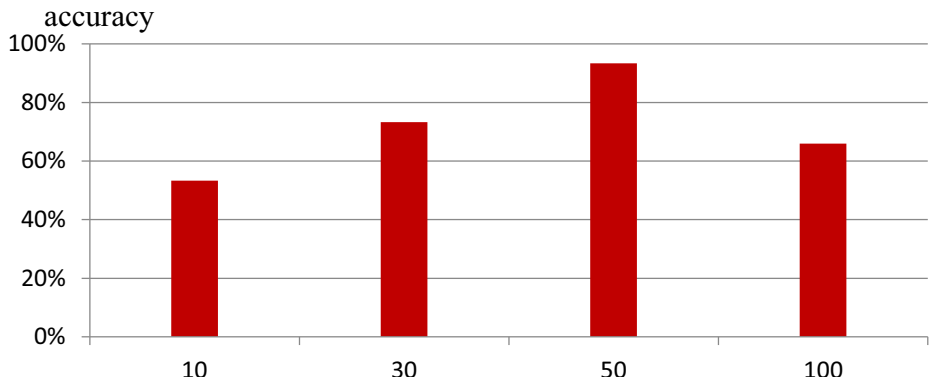
Table 3.3. Description of ELM Architecture

Annotations	Total(*node*)
Input	900
Hidden	50
Output	3

The architecture that will be used in this network consists of 3 layers those are input layer, hidden layer, and output layer. Input data are 900 nodes, hidden are 50 nodes and output consists of 3 nodes (Primary, Billion and Normal). Hidden nodes are determined randomly (Huang, 2006). It can be determined through several trials of system requirements. The selection of hidden nodes totaling 50 nodes is a good weight in producing high accuracy in tuberculosis disease classification systems and does not take much time in the image processing. Data to be entered in the input layer will be transformed

first. The training is carried out to find the optimal and appropriate weights and biases to be used in the testing process.

The test is carried out with *EPOCH* value = 1000, Learning Rate = 0.2 and different hidden nodes, starting from 10, 30, 50 and 100. The results of the test can be seen in Figure 4.10. Based on the results of each testing process with different Hidden node values, the Hidden node value will be able to classify tuberculosis with a high degree of accuracy.



Based on data from test results that have been carried out on tuberculosis disease classification applications from X-ray images of human lungs using extra learning machines, can be obtained the value of accuracy in identifying tuberculosis with an average of %.

$$\boldsymbol{Percentage\ of\ Accuracy} = \frac{\boldsymbol{Number\ of\ correct\ test\ data\ image}}{\boldsymbol{Overall\ amount\ of\ test\ data\ image}} \times \mathbf{100\%}$$

$$= \frac{\mathbf{14}}{\mathbf{15}} \times \mathbf{100\%}$$

$$= \mathbf{93.33\%}$$

From calculations above it can be seen that the accuracy of the method of extreme learning machine in classifying tuberculosis from X-ray images of human lungs can reach 93.33%.

4. Conclusion

It can be concluded based on the results of testing the tuberculosis disease identification system using Extreme Learning Machine that:

1. Extreme Learning Machine (ELM) method is able to classify tuberculosis through a good image. So that the results of the classification process of tuberculosis through lung X-ray images have an accuracy rate 93.33%.
2. Hidden Node greatly affects accuracy. After going through a number of tests, the smaller the hidden node value, the smaller the accuracy level obtained, otherwise the greater the Hidden Node value, the greater the accuracy level obtained. The value of the Hidden node ≥ 30 is a good parameter for classifying brain hemorrhage using Extreme Learning Machine.
3. In the image processing, it is necessary to determine the appropriate image *Threshold* value because it will have a significant effect if the used *Threshold* value is not appropriate then it will affect the system accuracy.

Acknowledgment

This research involves collaboration between students and lecturers in the Fasilkom -TI Universitas Sumatera Utara

References

[1]. Alsagaff, Hood dan Mukti, Abdul. 2010. "*Dasar-Dasar Ilmu Penyakit Paru"*. Airlangga University Press. Surabaya Andrian, Yudhi. Perbandingan metode LSB, LSB+1, dan MSB pada steganografi citra digital. STMIK potensi Utama. Medan.

[2]. Arie Yulianto, dr. 2014. *Tuberkulosis Paru: Penyebab Kematian Ke-2 Di Indonesia*. Kesehatan Umum. http://www.tanyadok.com/kesehatan/tuberkulosis-paru-penyebab-kematian-ke-2-di-indonesia.

[3]. Chen, F.-L. & Ou, T.-Y., 2011. Constructing a Sales Forecasting Model by Integrating GRA and ELM: A Case Study for Retail Industry. *International Journal of Electronic Business Management,* pp. 107-121.

[4]. Efford, N., 2000. Chapter 11 "Morphological image processing". In Efford, N. *Digital Image Processing: A Practical Introduction Using JavaTM*. Pearson Education

[5]. Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. 2006. Extreme learning machine : theory and applications. Int. J. of Neurocomputing 70(2006): 489-501.

[6]. Junifer, Jefri. 2016. "Mendiagnosis Penyakit Diabetes Melitus Dengan Menggunakan Metode *Extreme Learning Machine.* Jurnal ISD. Vol. 2 No. 2.

[7]. Kadir, Abdul. 2013. *Teori dan Aplikasi Pengolahan Citra*, Yogyakarta: Andi

[8]. Law, M. Andrew 2015. "Relative Optical Navvigation Around Small Bodies Via Extreme Learning Machines. United States, University Of arizona.

[9]. http://repositori.usu.ac.id/handle/123456789/8815

[10]. Library ECG. 2017 , http://www.ecglibrary.com.

[11]. Parker, J.R. 1994. *Pratical Computer Using C*. United State of America: John Willey & Sons, Inc.

[12]. Murdoko and Saparudin. 2015 "Klasifikasi Citra Daun Tanaman Menggunakan Metode Extreme Learning Machine" Konferensi Nasional Informatika (KNIF).

[13]. Nurminen, Antti. 1996. "Recording, Processing, and Archiving Carbonized Papyri." Diakses dari www.cs.hut.fi/papyrus/index.html.

[14]. Rahmadewi, Reni. 2016. "Klasifikasi penyakit paru berdasarkan citra rontgen dengan metode segmentasi sobel, Jurnal nasional teknik elektro,"vol: 5, No. 1.

[15]. Saputra, Andri. 2011. "*Sistem Pakar Identifikasi Penyakit Paru-Paru Pada Manusia Menggunakan Pemrograman Visual Basic 6.0".* Jurnal Teknologi Dan Informatika (Teknomatika). Vol. 1 No. 3.

[16]. Setiarini, Asih. 2017. "Identifikasi Bakteri pada Citra Dahak Penderita Tubercolusis (TBC) Menggunakan Metode Watershed". Journal of electrical electronic control and automotive engineering (JEECAE). Vol 2. No. 1.

[17]. Shahsavari. 2016. "Efficient Classification of Parkinson's Disease Using Extreme Learning Machine and Hybrid Particle Swarm Optimization". 4th International Conference on Control, Instrumentation, and Automation (ICCIA) 27-28. Qazvin Islamic Azad University, Qazvin, Iran.

[18]. Sumrina, Kanwal. 2016. "Towards a Novel Medical Diagnosis System For Cllinical Decision Support System Application. Scotland, UK. Uneversity os Stirling.

[19]. Sunyoto, Andi. 2013. "Analisa Metode Moment Invariant Untuk Mendeteksi Obyek Yang Telah Mengalami Transformasi". Jurnal Dasi. Vol. 14, No. 1. STMIK AMIKOM. Yogyakarta

[20]. Supriyati, Endang dan Khotimah, Tutik. 2014. " Deteksi Iris Mata Untuk Menentukan Kelebihan Kolesterol Menggunakan Ekstrasi Ciri Moment Invariant Dengan K-*Means*

clustering". Prosiding SNATIF ke 1. Program Studi Teknik Informatika, Universitas Maria Kudus.

[21]. W. K. Pratt. 1991. *Digital Image Processing,* 2nd ed., Wiley-Interscience, New York.

[22]. Xiang, Junlong, Westerland, Magnus. 2015. "Using ELM For Intrusion Detection In a Big Data Environtment. German. University Of Helsinky.