

Identification of Lung Cancer Using a Back Propagation Neural Network

Bambang Guruh Irianto*, M. Ridha Mak'ruf, Dyah Titisari

Department of Electromedical Engineering Politeknik Kesehatan Kemenkes Surabaya, Indonesia

Jl Pucang Jajar Timur 10 Surabaya, 031-5037095

*Corresponding author, e-mail: bgi_ps@yahoo.com

Abstract

Reading image of lung cancer screening well-known as X-ray by practitioners are sometimes subjective. This research tried to create software that can detect lung cancer as a comparison of the work of medical practitioners using artificial neural networks (ANN), with X-ray movies taken from the tool diagnostic radiography (DR) stored in the compact disc. The dependent variable observation in this study is the identification of DR X-ray image size of 1024 x 1024 pixels. A total of 10 images X-ray which has been observed by the physician radiology, with 5 images X-ray normal and 5 images lung cancer. In this study, the image processing is done through three stages: neighborhood averaging, median filter and histogram equalization. The results of these features are grouped in normal categories. From test results stating the truth is 80%. To facilitate the user in the lung disease pattern recognition. GUI applications design using MATLAB. We use some form of image processing which includes form training and testing. The best parameters obtained from this research include learning rate=0.3, the number of hidden layer=30 and tolerance error= 10^{-8} . From the results obtained by the level of accuracy of the training image of normal lung, lung cancer in a row is 80%. Overall the level of accuracy of the results is 80%.

Keywords: neural network, lung cancer, diagnostic radiography

Copyright © 2015 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Lung cancer is the world's health problems. From year to year, statistical data in various countries showed lung cancer incidence rates have tended to increase. Cigarette smoking is the main cause of about 90% of lung cancer cases in men and about 70% in women. The more cigarettes smoked, the greater the risk to suffer from lung cancer.

One examination of lung cancer is to use checks or better known as radiology x-rays (x-rays). The working principle of this instrument is based on x-ray diffraction. Introduction to the simple x-ray is the most commonly used techniques. The image of the x-ray will give different results between the lungs of healthy and unhealthy, such as lung cancer and stage of the cancer of the lung.

However, examination of lung cancer from x-rays result image still has a shortfall of some medical practitioners such as physicians of lung specialists still rely on visual observations in the reading results of the x-rays so that the result is very subjective.

Lung specialist doctor should perform x-rays image observations carefully and diagnosis that are completely accurate in the detection of lung cancer in patients. Therefore required a software capable of detecting lung cancer as a comparison of the medical practitioners, so that the software can help the accuracy of determining the identification of lung cancer.

Given the need for pattern recognition system encourage applications software as a system that can help humans perform activities. In this case is to carry out identification using neural networks. This research which uses neural networks has started many done like: A New Neural Network Back Propagation Algorithm and Its Application in English Education Evaluation [1]. The paper improves the original Neural Network Back Propagation with immune genetic algorithm to speed up its calculations and presents a new convergence of Neural Network Back Propagation algorithm for evaluating system performance in English education, A Novel Method to Optimize the Structure of BP Neural Networks [2]. The simulation experiment shows that this

method can select a more suitable network structure, and can ensure the minimum output error with the optimum structure of the network.

With the development of Neural Network utilization, then it can be used to pattern recognition network on the human body that has many variations. With this application are also applied to the world of health. This is evidenced by the existence of research such as: Classification of colon cancer using the method of GLCM and ANN Back propagation [3], Comparisons of the results of the classification of disease diabetes mellitus using Neural Network Back Propagation and learning vector quantization[4], Identifying glukoma eye disease based on the output of the visual field analyzer with genetic algorithm-neural network [5], Identification of breast cancer using back propagationneural network [6], The application of the Artificial Neural Network Back Propagation Method in the detection of Abnormalities of brain Ischemic Cerebral Infraction by the Delphi programming language [7], Clinical decision support system by using artificial neural networks to detect stage lung cancer sufferers carcinoma not small cell [8], Classification of lung Image with Histogram Feature Extraction and Neural Network Back Propagation [9], lung cancer detection From images of X-rays using Neural Network Back Propagation [10]

2. The Purposed of the Identification of Lung Cancer using Neural Network Back Propagation

Research on the utilization of neural network used in the radiology using X ray film, but still suffered flaws that allowed some of the information contained in the film lost a lot X ray film, so it can't be read by Back Propagation Neural Network. Therefore in this study, to identify lung cancer is taken the result image Diagnostic Radiography (DR) stored in Compact Disc. For this diagnosis is based by the experience of a doctor, so the learning process for beginner readers x ray image a bit difficult because the reader beginner requires accuracy and experience it yourself. Therefore, the researchers took the title identification of lung cancer using Neural Network Back Propagation.

2.1. Formulation Of The Problem

- 1) The extent to which the level of accuracy in identifying the neural network software image sourced from the Diagnostic Radiography (DR) stored in a CD (compact disc)?
- 2) How accurate Neural Network Back Propagation Software in identifying pictures of normal lung and lung cancer?

2.2. Research Purposes

General Purpose

Designing an application Neural Network Back Propagation program Software to identify Lung Cancer.

Special Purpose:

- 1) To identify the x-rays in patients with lung cancer by a physician.
- 2) To make the application Neural Network Back Propagation program Software
- 3) To train a Neural Network Back Propagation Software for the introduction of a normal lung.
- 4) To apply the design of artificial neural network software for image Diagnostic Radiography (DR) of lung cancer.

2.3. Benefits

To add to the Knowledge about the utilization of ANN in identifying lung cancer using film from Digital Radiography.

2.4. Conceptual Framework

Data taken in this research is a form of x-ray images result Digital Radiography (DR) of photographs of the lungs which then stored in a Compact Disc [11, 12]. The format of the data .Jpeg is in normal patients and lung cancer data. Furthermore, the image processing (image processing) through three stages, namely using averaging neighborhood is one of the low-pass filter, which works by replacing the value of a pixel in the original image with the average value

of the pixel and its neighboring environment. Photos of the original size of 1024 x 1024 pixels is converted into 320 x 240 pixels in order to display the software is not too large. From the photo 320 x 240 pixel size used as a matrix of 20 x 15 as adjusted by the computer in the process of identifying the neural network. Furthermore, the image processing comprising: Neighborhood Averaging, histogram equalization and median filter to soften the image or images and earned value metrics are then used as input of neural network through the training phase, the process of testing and identification process [13].

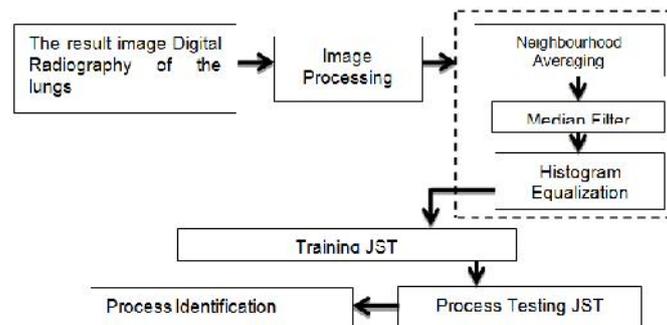


Figure 1. Conceptual Framework

The process of testing is done by inserting normal lung image data and lung cancer picture. Then the calculate activation, hidden and output of the entire image data included in the process of testing (testing) will recognize the image of normal lung or cancer.

3. Research Metodology

This type of research is research pre experimental (after only design). On the design of these researchers only see results without measuring the previous state. The control group in this study is the images from Digital Radiography results. In this research is to create an application program (software) that Neural Networks are applied in the result image Digital Radiography [14, 15].

Independent variables in this study are the process of training (training) and the process of testing (testing) is performed. Dependent variables in this study are the identification by the observation results. Of the entire image data entered will be known level of accuracy in recognizing software image data entered. Object of research: Image Lungs results Digital Radiography (DR) at *Department of Electromedical Engineering Poltekkes Kemenkes Surabaya*. Target research: Figure lungs Digital Radiography results of normal and abnormal are stored in the form of a Compact Disc (CD) with a size of 1024 x 1024 pixels. Location and Time Research: Location of research done at *Department of Electromedical Engineering Poltekkes Kemenkes Surabaya* and research time for 6 months, starting from June to November 2014. Data processing and data analysis: To test the credibility of the data is done by Triangulation Technique i.e. by checking data on the same source with different techniques, the data obtained in *Department Electromedical Engineering Poltekkes Kemenkes Surabaya* of the result image Digital Radiography stored in Compact Disc (CD). After the data is obtained from the field, then do image processing for ease in doing the next process that is done training (learning) and testing. By using the method of BPNeural Network will be able to Identify the normal lung image and lung cancer image.

4. Results and Discussion: In this Study Image Processing is Done through 3 Stages. IE:

4.1. Neighborhood Averaging

In principle, the filters used in the neighborhood is one of a kind of averaging low-pass filter, which works by changing the value of a pixel in the original image with the average value of the pixel and its neighboring environment.

Listing Programme for Neighbourhood, Averaging is as follows:

```
% Neighborhood Averagin , kernel=[1 1 1;1 1 1;1 1 1]/9;
F=uint8(conv2(double(img),kernel,'same'));imwrite (F,'Neighborhood Averagin.bmp');
```

4.2. Median Filter

The median filter is a type of low-pass filter, which works by changing the value of a pixel in the original image with the median value of the pixel and its neighboring environment. Compared to the neighborhood, this averaging filter more insensitive to extreme intensity differences.

Program listings for the median filter is as follows:

```
% FILTER MEDIAN; D=medfilt2 (F); imwrite (D,'filtermedian.bmp');
```

4.3. Histogram Equalization

Histogram Equalization process program listings:

```
%EKUALISI HISTOGRAM ADAPTIVE; E=adaphstheq (D);imwrite
(E,'ekualisihistogram.bmp');
```

Histogram Equalization process then it will appear the following graph:

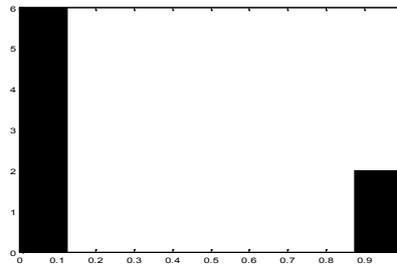


Figure 2. Graph Image Histogram Equalization process passes through the lung. The matrix of the value that will be entered as input BP Neural Network

4.4. Process Training

The training data is done with the variation hidden layer and the number of epoch in order to get the most performance network architecture approach the target error 0.001. This variation of the results obtained from the parameters used in the process of training namely hidden layer=10, epoch=3000, learning rate=0.1 an target error=0.001. Training Neural Network used in the study are using the method of Neural Network Back Propagation. The image of the used training amounted to 10 image, consisting of 5 normal lung image, and 5 image of lung cancer. The results showed that the training of 10 image training consisting of 5 normal image and 5 image of lung cancer acquired training of 100% accuracy. Where each image is classified correctly based on the target of each category, can be seen in the table above.

4.5. Testing Process

Testing was done using input parameters obtained in the process of training the learning rate of 0.3, neurons in the hidden layer numbering 30, maximum error tolerance is 10^{-8} and a maximum of epoch as much as 1000 times. The target used in the testing in accordance with the targets that were used in the training of 10 image training consisting of 5 normal image and 5 image of lung cancer. The results indicate that there are 2 images that have different output values of output with the desired target, namely the image of normal lung 1 and 1 piece of lung cancer imagery.

One of the factors causing the occurrence of irregularities were the result of extracting features that are available on the other image categories range. On the image of N_1 and N_2, the test results provide the results the presence of cancer in the lung. Features which have difference is its homogeneity, the value of its homogeneity image N_1 and N_2 is range i.e.

cancer imagery 0.94-0.95. Small range of its homogeneity image of lung cancer resulting in such features in the group in the category image of lung cancer.

4.6. Results Display Applications

To facilitate users in performing pulmonary disease pattern recognition. MATLAB GUI application uses design, with some forms: Form image processing, Form of training and testing. Following is the display of the results of each Form submission:



Figure 3. Display image processing Form

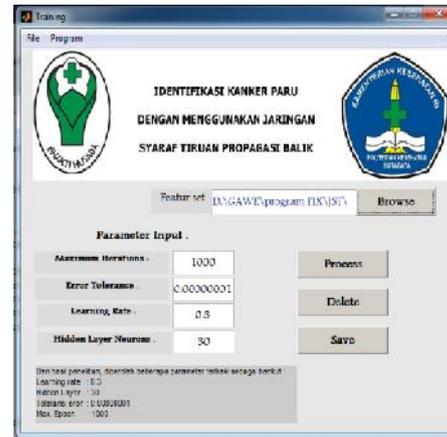


Figure 4. Display Form Training NN

Form image processing functions to display the image processing results so obtained the image results in the Form of testing. To display the image of the user simply presses the button 'show image' at the bottom right of the application.

Form of training is used to conduct training. Users only need to enter fiturset of imagery training and parameter values to be used. The training parameters, among others the maximum iterations, error tolerance, the learning rate and the number of neurons in the hidden layer. The user can press the button "delete" to replicate training and 'save' button to save the final weights of the training.

4.7. Display Form Testing Neural Network

Form of testing used to do pattern recognition pulmonary disease, users only need to enter image that will be identified, and by pressing the button 'Process' then the result will be displayed in a column of the results at the bottom left of the application.



Figure 5. Display the Normal image Testing



Figure 6. Display Image Cancer Testing

In the picture the display form data by inserting a test picture that will be tested, when we press the button, the process will appear to the test results, diagnoses of the image will appear are whether normal or pulmonary cancer.

In the example image above is one example of a normal image testing, diagnostic results obtained are normal.

Next put the data image of lung cancer. Display of the results obtained, shown in the picture below:

One example of the process of testing on images of lung cancer diagnosis results obtained are lung cancer. In the Form of testing there were several keys that are used to perform a pulmonary disease pattern recognition, namely:

The button 'Search' function to open the File that, *.Process ' button serves to manipulate the process testing. The user only need to enter image that will be identified, the button ' delete ' function to repeat the testing process by inserting a new image, * Menu ' Program ' was created to allow a user to open another Form.

4.8. The Discussion of the Accuracy of the Research Results

Table 1. Normal image and Lung cancer image

Category	Number of Image	Number of True	Number of False	Accuracy (%)
Normal Image	5	3	2	60
Lung Cancer Image	5	5	0	100
Total	10	8	2	80

After going through several stages in this study, from the start of Image Processing until the process of training and the process of testing, by inserting the image of testing a total of 10 images with details of Normal Lung Image 5 and 5 image of lung cancer. From the test results that are obtained from 5 Normal Lung Image, 4 Image can be recognized as a normal lung and 1 normal image known as the lung cancer rate of accuracy of 80%. And while the Image of lung cancer can be recognized as lung cancer Images 4 and 1 image of lung cancer known as the normal lung. So the overall testing process of data i.e. 10 data consisting of 5 normal lung image and 5 image of lung cancer, the research results obtained accuracy of 80%.

5. Conclusion

Based on the results of the study it can be concluded that: 1) The best Parameters obtained from this research include learning rate=0.3, the number of hidden layer=30 and error tolerance= 10^{-8} . 2) Of research results obtained accuracy of pulmonary imagery training normal, lung cancer respectively was 80%. 3) Overall the level of accuracy of the study results is of 80%.

References

- [1] Yu Zhike. A New Neural Network Back Propagation Algorithm and Its Application in English Education Evaluation, *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(11): 6441-6446.
- [2] Changming Qiau, Shuli Sun. A Novel Method to Optimize the Structure of BP Neural Networks, *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(10): 558-5593.
- [3] Agung Raditya Putra, Achmad Rizal, M Syahrul Mubarak. Klasifikasi Kanker Usus Besar Menggunakan Metode GLCM dan JST Neural Network Back Propagation (classification of colon cancer using the method of GLCM and ANN Neural Network Back Propagation). *Jurnal Elektro*. 2012; 5(2): 125-132.
- [4] Agus Nurkhozin, Mohammas Isa Irawan, Imam Mukhlas. *Komparasi Hasil Klasifikasi Penyakit Diabetes Melitus Menggunakan Jaringan Syaraf Tiruan Neural Network Back Propagation Dan Learning Vector Quantization (Comparations of the results of the classification of disease diabetes mellitus using Neural Network Back Propagation Neural network and learning vector quantization)*. Seminar Nasional MIPA UNY. 2011.

- [5] Agus Hartono. Identifying glukoma eye disease based on the output of the visual field analyzer with genetic algorithm-neural network. Teknik Fisika ITS. Surabaya. 2003
- [6] Rizfa Zulfika. Identification of breast cancer using neural network back propagation. Surabaya, Teknik Fisika ITS; 2006.
- [7] Nurhayati. The application of the Neural Network Back Propagation Method in the detection of Abnormalities of brain Ischemic Cerebral Infraction by the Delphi programming language. *Jurnal Fisika dan Aplikasinya, ITS*. 2010; 6(1).
- [8] R Kurniawan, S Hartati. Clinical decision support system by using neural networks to detect stage lung cancer sufferers kind carsinoma not small cell. *IJCCS*. 2011; 5(2): 35-45.
- [9] Hasan Bisri, M Arief Bustomi, Endah Purwanti. Classification of lung Image with Histogram Feature Extraction and Neural Network Back Propagation Neural Network. *Jurnal Sains dan Seni Pomits*. 2013; 2(2): 68-71.
- [10] Tri Deviasari Wulan, Endah Purwanti, Moh Yasin. Lung cancer detection from the image of x rays using back propagationneural network. *Jurnal Fisika dan Terapannya*. 2013; 1(1): 58-64.
- [11] Radiography Working Principle. 2014.
- [12] Rasyid, Dwi Rochmayanti. Modul Teknik Pesawat Radiologi. Poltekkes Semarang. 15.
- [13] Bambang Yuwono. Image smoothing using Mean Filtering, Median Filtering, Modus Filtering and Gaussian Filtering. *Jurnal Telematika*. 2010; 17(1).
- [14] Notoatmojo Soekidjo. Health reseach methodology. Cetakan Kedua. Jakarta: PT Rineka Cipta. 2002.
- [15] Sugiyono. Statistika untuk Penelitian. Cetakan Ketiga belas. Bandung: CV ALFABETA. 2008.