ISBN: 978-602-5649-417

DETECTION ABNORMALITIES OF THE X-RAY IMAGE OF THE LUNG BASED ON THE GRAY LEVEL COOCCURANCE MATRIX USING NEURAL NETWORK

Tri Deviasari Wulan[1]

[1] Department of Information Systems, Faculty of Engineering, Universitas Nahdlatul Ulama Surabaya Surabaya, Indonesia
tridevi@unusa.ac.id

Abstract

In this study a software system has been built that can detect abnormalities of x-ray images using backpropagation neural network. This software is expected to be a tool for lung doctors to diagnose lung cancer quickly and accurately. the results of this software are normal x-ray images as well as those diagnosed with lung cancer. this process begins with the preprocessing stage, the second stage is feature extraction using Gray level co-occurance matrix (GLCM) to look for feature texture characteristics of the pulmonary x-ray image. the third stage is classification using backpropagation neural network. Classification using backpropagation consists of the training and testing stages. The parameters used for the training and testing stages are learning rate = 0.1 and target error = 0.001. at the training stage, there are various hidden layer and epoch variations to find the best network architecture. The training results show hidden layer = 15 and the number of epoch = 200 is the best neural network architecture with an MSE value of 0.165. the testing phase using the same parameters shows the accuracy of 87.5%

Keywords: ***lung cancer, GLCM, backpropagation***

Introduction

Lung cancer is the leading cause of death in the world because it reaches 30 percents of all types of cancer. In men, lung cancer accounts for 1/3 of all deaths caused by cancer. Based on WHO data, lung cancer is the highest cause of death in men (21.8%) compared to other types of cancer, namely prostate, liver, and nasopharynx cancer. Whereas in women, lung cancer is the second leading cause of death after breast cancer. 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 for lung cancer. Based on data from the Department of Pulmonology and Medical Respiration FKUI-RSUP Persahabatan, the number of new cases of lung cancer increased more than 5 times in the last 10 years and most of the patients came at an advanced stage (IIIB / IV). (KPKN, 2017)

The first diagnosis of lung cancer is to use an X-ray machine. This machine is widely used because the price is relatively cheap and most hospitals have this machine. However, examination of lung cancer using an X-ray machine still has its drawbacks. Some media practitioners such as pulmonary doctors are still very subjective in reading X-rays. The doctor's observation of the results of X-rays must be done carefully so that the diagnosis of the patient is accurate. In addition, the diagnosis of lung cancer at an early stage will provide a higher chance of patient life. Therefore we need a software that is able to detect lung cancer as a comparison of the work of lung doctors. So that this software can help accuracy in diagnosing lung cancer.

The study was conducted by Huan Wang, et al in 2010 about the Multi Binomonal Logistic Prediction model in malignant lung nodules based on the feature textures in CT Scan images. In this study nodule segmentation is done by searching for Region of interest (ROI) pulmonary nodule then extracted using the feature texture method, namely gray level co-occurrence matrix (GLCM). Although the results reported are accurate, this study still uses statistical analysis for the method of classification analysis between benign and malignant pulmonary nodules.Research has also been carried out by Azian Azamimi Abdullah et al. in 2012 by designing a CAD (Computer Aided Diagnosis) using artificial neural networks to detect and classify lung cancer in X-Ray images. Artificial imitation network used in this research is backpropagation. From these studies, it was found that backpropagation artificial neural networks have been successfully used to develop CAD in detecting and classifying lung cancer from X-Ray images.

This research has designed a software system to detect the abnormalities of pulmonary X-Ray Thorax images based on GLCM using backpropagation neural network. Chest X-ray of human X-rays used aged 35 years and over is caused by this age range most affected by lung cancer. From the results of the photo data, digital image processing is carried out before the training process is carried out using artificial neural networks. This software is made using backpropagation artificial neural networks, which is one of the algorithms in artificial neural networks that use multilayer, because the more layers used are expected to produce more accurate results (Desiani, 2007). The output of the lung cancer detection system is abnormal information from X-ray images.

Research methods

In general, the research methodology for the detection of abnormalities of x-ray images is shown in Figure 1.

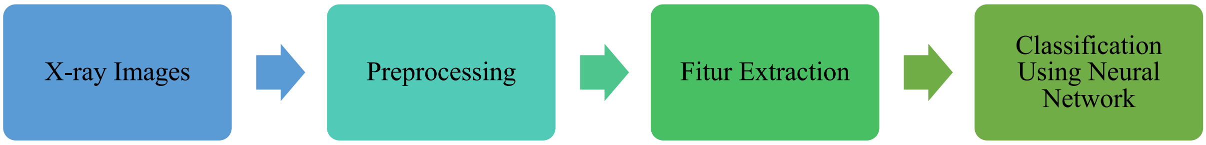


Figure 1. Research Methodology

The image used in this study is X-ray image consisting of x-ray images of lung cancer and normal. The first stage is image preprocessing which consists of 2 stages are cropping and resizing. the cropping stage aims to cut the image on the lung area and the resizing stage is used to change the image size to a smaller size to facilitate the subsequent image processing. The second stage is feature extraction which aims to get special features of pulmonary x-ray images.Feature extraction used in this study is feature texture extraction with the gray level co-occurrence matrix (GLCM) method. GLCM uses a second-order texture calculation taking into account the relationship between two-pair pairs of original images taken into account . Co-occurence is defined as the combined distribution of grayscale levels of two pixels which are separated by a certain distance and direction (Δx, Δy). Texture feature process is used to get different features from the x-ray image of lung cancer and normal. The features that are used as input in the classification consist of 3 are:

1. Contrast

   Contrast is a measure of the existence of gray pixel image variations, can be calculated using the equation:

$$Contrast = \sum_{n=1}^{L} n^{2} \left\{ \sum_{|i-j|=n} GLCM(i,j) \right\}$$

2. Correlation

   Correlation is a measure of the linear dependence of the degree of grayness of the image so that it can provide an indication of a linear structure in the image, can be searched using the equation:

$$Correlation = \frac{\sum_{i=1}^{L} \sum_{j=1}^{L} (ij)(GLCM(i,j) - \mu_i' \mu_j'}{\sigma_i' \sigma_j'}$$

3. Entropy

   Entropy shows the size of the shape irregularity. Large entropy values for images with evenly distributed gray grades and small values if the structure of the image is irregular or varied. This entropy value can be calculated using the equation:

$$Entropy = -\sum_{i=1}^{L} \sum_{j=1}^{L} (GLCM(i,j) \log (GLCM(i,j)$$

The third stage is the classification stage. In this study classification uses backpropagation neural network which consists of two processes are training and testing. The training process using backpropagation is shown in Figure 2. The testing process used are last weights from training process.

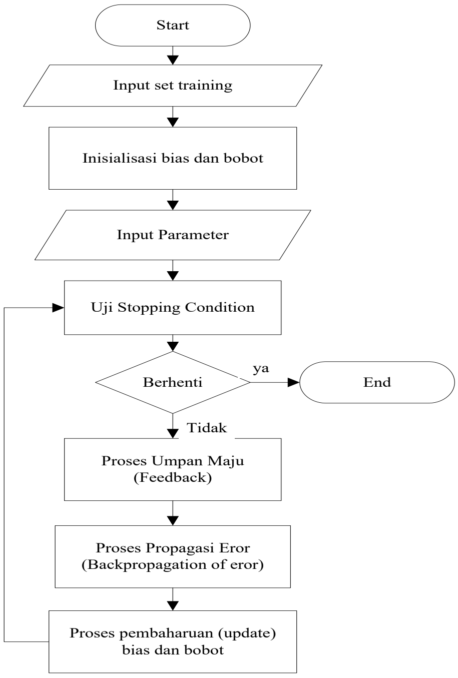
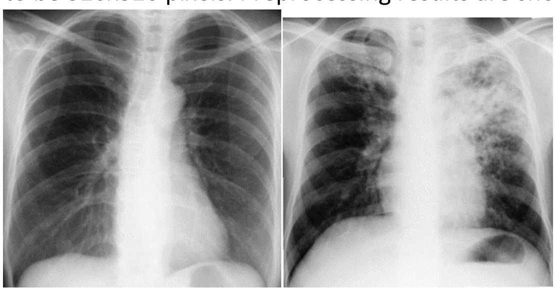


Figure 2. Training process of backpropagation neural network

Research Results and Discussion

1. Preprocessing x-ray images

Preprocessing x-ray images consists of 2 stages are cropping and resizing. The croping stage is done to get the lung area only and eliminate other areas. The next process is resizing to equalize the image after the cropping process, so the image size is set to be 320x320 pixels. Preprocessing results are shown in Figure 3.



(a) (b)

Figure 3. Preprocessing X-Ray Images.(a) Normal X-Ray Image (b) X-Ray Image of lung cancer

2. Feature Extraction

Feature texture extraction using GLCM aims to obtain feature differences between cancer and normal x-ray images. there are 3 GLCM features used in this research namely contrast, entropy and correlation. Table 1 shows the differences results in feature texture between lung cancer and the normal x-ray images of the three GLCM features used.

Tabel 1. the differences results in GLCM between lung cancer and the normal x-ray images.

Lung Cancer			
	Kontras	Entropi	Korelasi
GLCM 0	18,207	7,828	0.0003
GLCM 45	27,000	7,957	0.0003
GLCM 90	21,280	7,867	0.0003
GLCM 135	26,879	7,951	0.0003
Normal			
	Kontras	Entropi	Korelasi
GLCM 0	35,958	8,256	0.0002

GLCM 45	47,384	8,353	0.0002
GLCM 90	37,550	8,268	0.0002
GLCM 135	48,095	8,358	0.0002

3. Classification

The classification uses backpropagation neural network which consists of 2 processes are training and testing. The training process uses 20 thorax images x-ray data consisting of 10 x-ray images of lung cancer and 10 normal x-ray images. Data training was done to vary the hidden layer and the number of epochs to get the best network architecture seen from the performance (MSE) results that were closest to the error target. The parameters used in this training are the learning rate of 0.1 and the target error of 0.001. The hidden layer variation consists of 5, 10, and 15. While the variation in the number of epoch consists of 100, 150 and 200. Figure 4 shows the graph between performance (MSE) and variations in the number of epoch with all hidden layer variations. The training results show hidden layer 15 with the number of epoch 200 shows the best performance with MSE = 0.1650 and accuracy = 100%. The best network architecture produces the final weight used as a weight in the backpropagation neural network testing process.

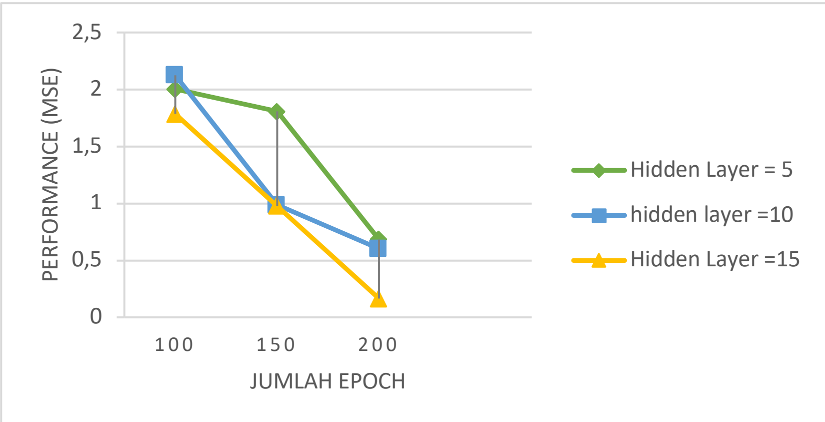


Figure 4. Graph between performance (MSE) and variations in the number of epoch with all hidden layer variations

The testing process used 8 chest x-ray images consisting of 4 x-ray images of lung cancer and 4 normal x-ray images. The weight used in backpropagation neural network testing is the result of the final weight in the training process. The parameters of backpropagation neural network testing are the same as those used in the training process. Table 2 shows the results of the system testing that has been carried out. From the results of system testing, the accuracy of the system is 87.5%.

Tabel2. the results of the system testing using backpropagation neural network

Input	Result	Target	Conclusion
Data 1	Kanker	Kanker	Sesuai
Data 2	Kanker	Kanker	Sesuai
Data 3	Kanker	Kanker	Sesuai
Data 4	Kanker	Kanker	Sesuai
Data 5	Normal	Normal	Sesuai
Data 6	Normal	Normal	Sesuai
Data 7	Normal	Normal	Sesuai
Data 8	Kanker	Normal	Tidak Sesuai

Conclusions and recommendations

Accuracy in the diagnosis of abnormalities of chest x-ray images is needed to improve the chances of healing and handling for patients. this system is made as a comparative work tool for pulmonary doctors in diagnosing the results of thorax x-ray images. There are 2 types of thorax x-ray images used, namely x-ray image of lung cancer and normal. Feature texture extraction is used to get the difference in features between cancer research and normal x-ray images. The GLCM feature uses contrast, correlation and entropy. The classification results using backpropagation neural network shows the network architecture with the number of epoch = 200 and hidden layer = 15 is the best architecture with a performance value (MSE) of 0.1650. The test results showed that the accuracy of the system in detecting abnormalities of the X-ray image of the thorax was 87.5%.

Further research can be developed by adding the amount of data used so that the level of accuracy can increase. This is because the amount of data greatly affects the accuracy of the neural network architecture used. In addition, using the Neural network method is also highly recommended to know the suitability of a data match with the method used. Another development that can be done is to add preprosessing stages that are done to further automate the area to be studied

References

Komite penanggulangan kanker nasional.2017. Pedoman Nasional Pelayanan Dokter Kanker Paru*. Kementrian kesehatan Indonesia*: Jakarta

Huan Wang., Guo, X., Jia.Z., Li, H, dkk. 2010. Multilevel Binominal Logistic Prediction Model For Malignant Pulmonary Nodules based on Texture Feature on CT Images. *European Journal of Radiologi* . pp 124-129.

Abdullah A.A & Shaharum S.M. 2012. Lung Cancer Cell Classification Method Using Artificial Neural Network.*Information Engineering Letter*.Vol 2.pp 49-58.

Desiani Anita. 2007. Kajian Pengenalan Wajah dengan Menggunakan Metode Face-Arg dan Jaringan Saraf Tiruan *Backpropagation*. *Media Informatika* Vol. 5 No. 2. Pp 99-111.