LUNG DISEASE IDENTIFICATION AND CLASSIFICATION THROUGH NEURAL NETWORKS

RHIA JOYCE O. EDEN[1], JOANNA MARIE A. ELLARINA[1], CHRISTINE R. MANRIQUE[1], RAFAEL JOHN B. SURNET[1] AND ORLAND D. TUBOLA[1,2]

[1]Department of Computer Engineering, College of Engineering, Polytechnic University of the Philippines
[2] Center for Engineering and Technology Research, Institute of Science and Technology Research, Polytechnic University of the Philippines

Abstract: Public hospitals lack attending doctors who can interpret X-ray films thus, making patients wait for long hours just to get the test results. This study uses Neural Networks in identifying and classifying lung diseases based on X-ray image samples. The network can make the process of recognizing the disease faster, and doctors can just verify the result that the system has given. The chest X-ray samples were gathered from different hospitals around Metro Manila. These image samples were resized and had undergone image processing. Ten distinct points were acquired from each image using the Principal Component Analysis. All the points obtained from the images were placed in a single mat file. This mat file acts as the input data for the training of the Neural Network in the MatLab software. There are two hidden layers and has 10 neurons for each. The network type used is Feed-forward, Levenberg-Marquardt back-propagation with trainlm as the training function. The adaption learning function and performance function used are learngdm and Mean Square Error (MSE), respectively. The Neural Network training state, performance and regression were shown. From the total of 206 inputs, 186 of those were correctly classified according to their disease. The Pleural Effusion has the highest accuracy with 94.87% and the Normal has the lowest with 79.31%.

***Keywords**: neural networks, lung disease, X-ray, principal component analysis, feed-forward, back-propagation, mean square error*

1. INTRODUCTION

Nowadays, lung diseases are one of the most serious threats to human health. Some phenomena such as air pollution, various infections, and smoking habit have recently increased its risk factors drastically. According to (Crapo, 1994), physicians conduct lung function tests with these procedures: evaluate symptoms and signs of lung disease, assess the progression of lung diseases, monitor the effectiveness of therapy, evaluate preoperative patients in selected situations, screen people at risk of pulmonary diseases such as smokers or people with occupational exposure to toxic substances in occupational survey and to monitor for the potentially toxic effects of certain drugs or chemicals. Lung function tests are often performed using spirometry. Spirometry is the measurement of the air moving in and out of the lungs during various respiratory maneuvers. It allows one to determine how much air can be inhaled and exhaled, and how fast (Gold, W.M., 2000).

An x-ray (radiograph) is a non-invasive medical test that helps physicians diagnose and treat medical conditions. The chest x-ray is the most commonly performed diagnostic x-ray examination. It produces images of the heart, lungs, airways, blood vessels and the bones of the spine and chest. Chest x-rays help doctors diagnose conditions such as pneumonia, heart failure, lung cancer, lung tissue scarring, and sarcoidosis. Medical experts conduct this test for early diagnosis of lung conditions which can help alleviate symptoms and slow the progression of the disease.

Artificial Neural Networks (ANN) is currently a hot research area in medicine and it is believed that it will receive extensive application to biomedical systems in the next few years. Neural networks are ideal in recognizing diseases using scans since there is no need to provide a specific algorithm on how to identify the disease.

Neural Networks techniques have recently been applied to many medical diagnosis problems. But there has not been a significant use in a hospital or clinic routinely. The reason is that people don't think machines to be much reliable when it comes to diagnosis of a disease. But, soft computing tools like NNs, Fuzzy Logic (FL), and Genetic Algorithm (GA), can do well to ease and complement the work of medical experts. They can help to filter out the real patients, which will reduce the costs and time required for diagnosis. The doctors can then provide all their attention to the actual patients.

In a recent study, they developed another research for the classification and detection of restrictive respiratory abnormalities by means of artificial neural networks. In this study, it was revealed that artificial neural networks could be used for the detection of normal and restrictive patterns of lung diseases.

There is currently considerable interest in developing methods for quantitative imaging of the lung to identify disease early and to follow disease progression or to evaluate the result of interventions. There is a hope that quantitative measures based on lung imaging will provide surrogate markers that have enhanced sensitivity and specificity relative to more conventional pulmonary function tests or other more subjective clinical parameters thus accelerating the evaluation of disease interventions.

2. METHODOLOGY

2.1 Gathering of X-ray samples

Samples were collected from different medical centers in Caloocan, Manila and Quezon City. Hospital institutions were very strict regarding their X-ray film database. From all the hospitals and medical centers that we inquired, Jose Reyes Memorial Medical Center, San Lazaro Hospital, Manila Central University – Filimon D. Tan Choco Medical Foundation and Caloocan City Medical Center kindly confirmed and were willing to share their records with us.

Every chest X-ray was placed in the projector for clear image capture. The soft copy of the chest X-ray images were then transferred to a computer. It was cropped to a size that the lungs, including some portions of the diaphragm of the patient was only shown in the image which served as the sample data.

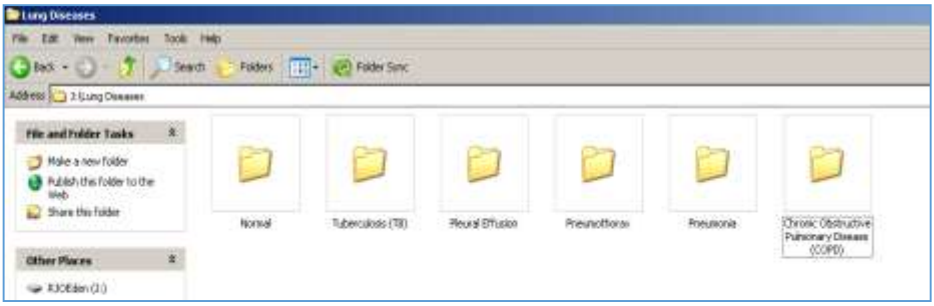

Figure 1. Database folder for different lung diseases

In this study, the researchers used 210 X-ray images of six different lung conditions (including the normal condition) collected from the stated hospitals in Metro Manila. Each condition has a minimum of 31 and a maximum of 39 X-ray images.

2.2 Image Processing Techniques

After the images were cropped, the images were then resized to 50x50 pixels. The lung X-ray image is converted into a gray scale intensity image. Threshold technique is then applied to the gray scaled image. This image analysis technique is a type of image segmentation that isolates objects by converting gray scale images into binary images. Thresholding an image is one way of partitioning an image into a foreground and background. Image thresholding is most effective in images with high levels of contrast.

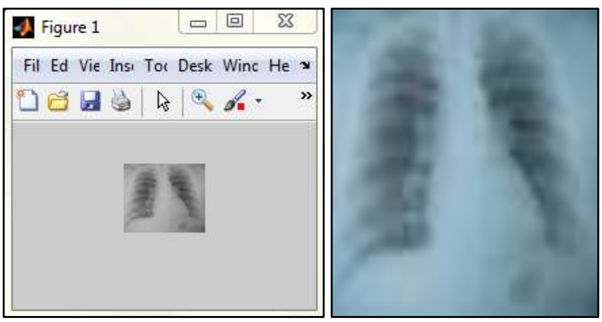

Figure 2. X-ray film of normal lung and its grayscale image

Figure 3. X-ray film of a lung with COPD and its binary image

Figures 2 and 3 shows ordinary X-ray sample of a normal lung and a lung with COPD with their processed image by gray scaling and converting it to a binary image. These image processing techniques make the recognition easier.

2.3 Neural Networks Training

The processed images are subjected to the algorithm which gets ten distinct points from it in order for the system to classify the disease that the X-ray image possessed. After all the images had their ten distinct points, making an array of values, it is then compiled to a mat file as an input data. On a separate mat file, every disease must have one column vector for the sample data. Another mat file must be created for the target data which consists of only one value per column. The value written in the target data depends on the assigned number of the classification of the input data. Finally, the data is ready for training. The nntool function in MatLab is used for the network training. The input and target data are imported. Then, a network consisting of the input and the target will be created and the feed-forward will be used, back-propagation as the network type. The adaption learning function and performance function used are learngdm and Mean Square Error (MSE), respectively. Afterwards, the network will be trained. Import the network created and use the network's name in the algorithm.

This network has two hidden layers and there are 10 neurons for every layer. Increasing the number of neurons in the hidden layer increases the power of the network, but requires more computation and is more likely to produce overfitting. The training function and transfer function used are trainlm and tansig, respectively. The accuracy of network depends upon the training and database. A minimum of 4 days were spent to train the whole network.

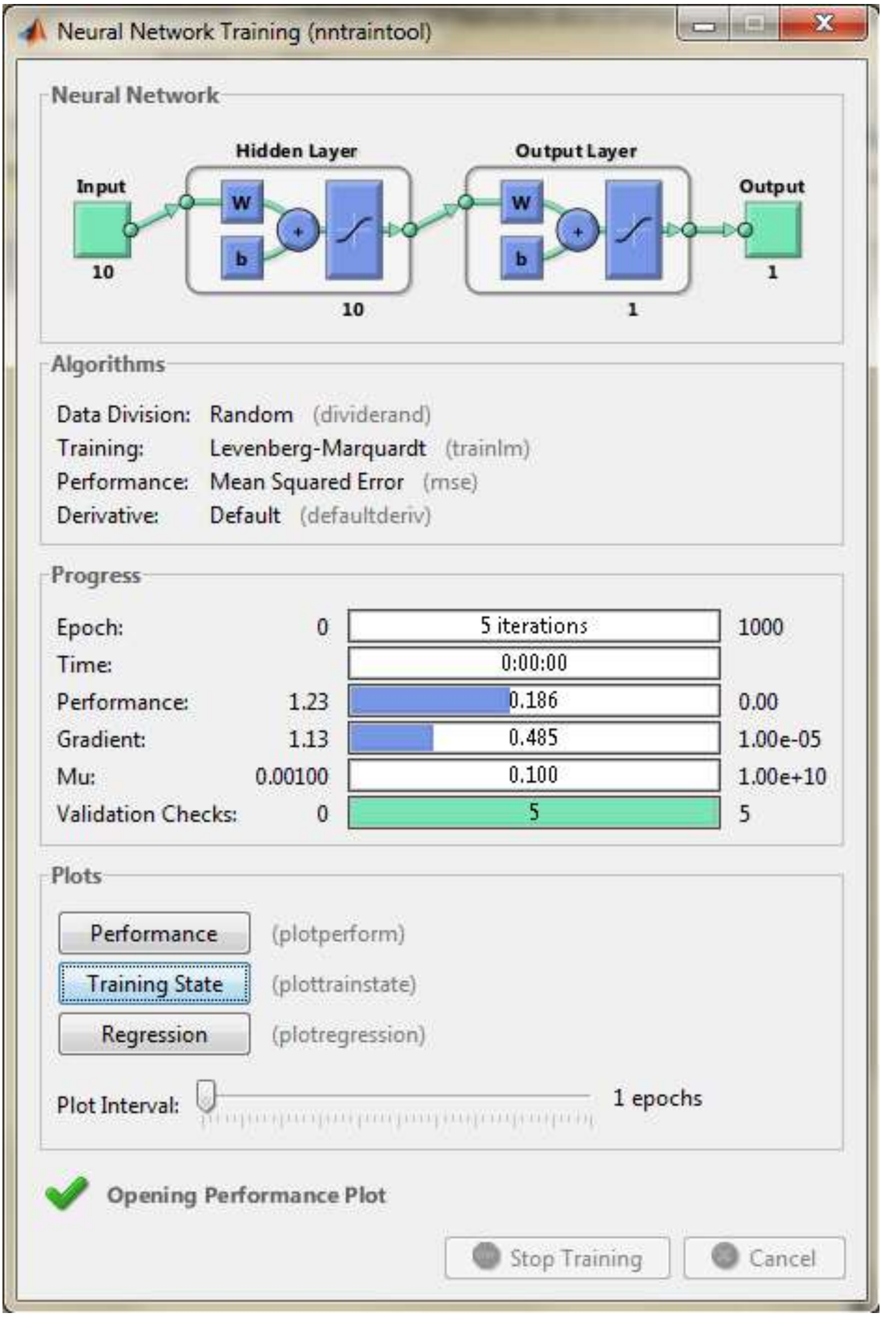

Figure 4. Neural Network Training

3. RESULTS AND DISCUSSION

After several training and testing, the proposed system for classifying and identifying lung diseases utilizing Neural Networks was complete. There were six different lung conditions including the Normal, Chronic Obstructive Pulmonary Disease, Pleural Effusion, Pneumonia, Pneumothorax and Tuberculosis that were recognized by the system.

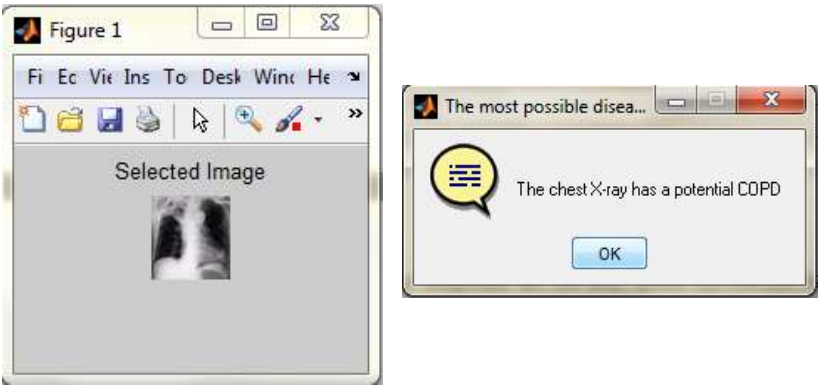

Figure 5. X-ray result of COPD

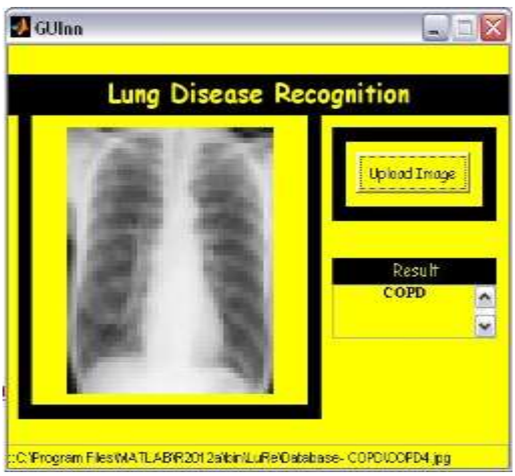

Figure 6. GUI Result of COPD X-ray

Figure 5 shows the result of a particular lung which has a potential COPD while Figure 6 showcases the enhanced GUI-based system of the Neural Network and displays the result of the identified and classified lung disease based on the X-ray image uploaded. The GUI was made in order for the user to use the system easily and with little supervision.

However, the system outputs some inaccurate results. There were some x-ray samples that were not correctly recognized by the system. Like a Normal x-ray sample was recognized as a Pleural Effusion X-ray which results the Normal with the lowest percentage of accuracy with 79.31%. Some Pneumothorax X-ray samples with an 88.57% percentage of accuracy were recognized as Tuberculosis. Some Pneumonia X-ray samples were recognized as Tuberculosis and COPD with an 89.19% percentage of accuracy. Some Pleural Effusion X-ray samples were recognized as a Pneumothorax and Tuberculosis while some COPD and Tuberculosis were recognized as a Pleural Effusion but both of these have an outstanding percentage of accuracy.

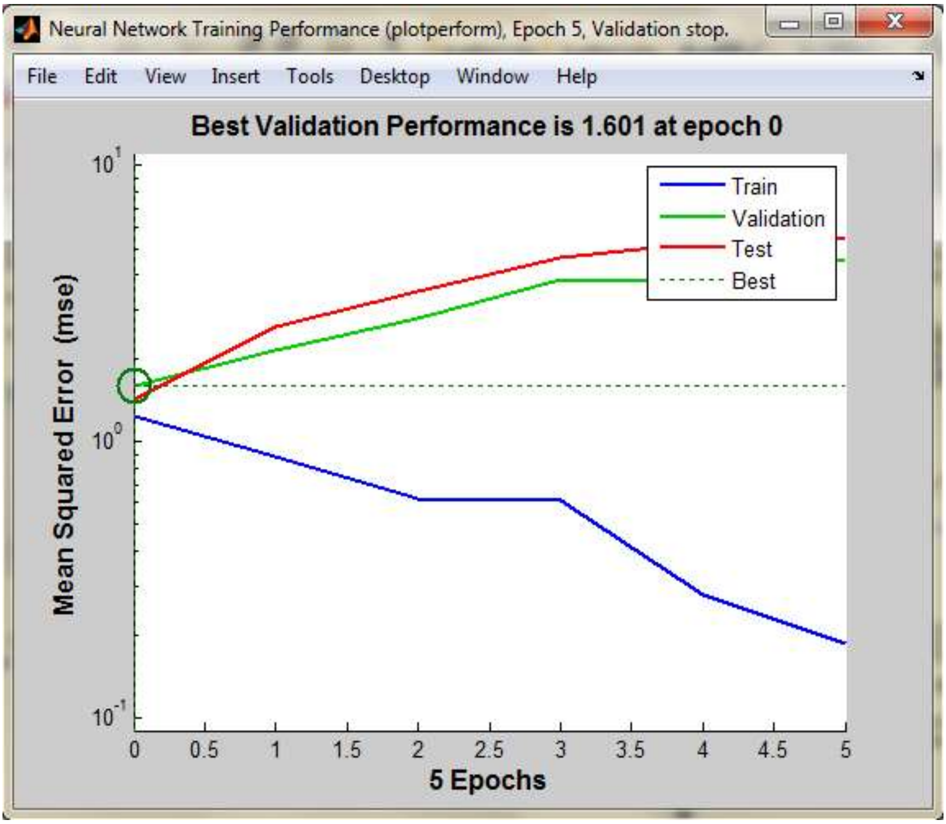

Figure 7. Neural Network Training Performance

Figure 7 shows the training performance of each X-ray samples. Here, you can initially conclude if your system works accurately.

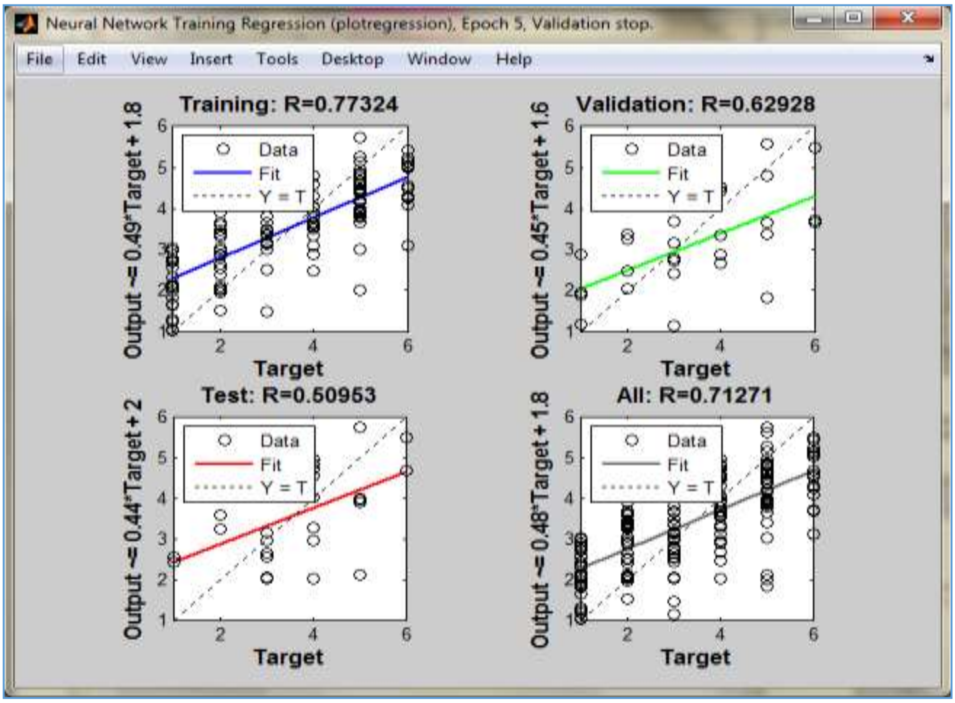

Figure 8. Neural Network Training Regression

Figure 8 shows the training regression of each X-ray sample. The three plots represent the training, validation, and testing data. The dashed line in each plot represents the perfect result – outputs = targets. The solid line represents the best fit linear regression line between outputs and targets. The R value is an indication of the relationship between the outputs and targets. If R = 1, this indicates that there is an exact linear relationship between outputs and targets. If R is close to zero, then there is no linear relationship between outputs and targets.

Table 1 Experimental Results for the Proposed System of Lung Diseases Recognition

Lung Diagnosis	Number of Input Data	Number of Input Data with Correct Classification	Percentage of Accurate Classification
Normal	29	23	79.31
COPD	35	33	94.29
Pleural Effusion	39	37	94.87
Pneumonia	37	33	89.19
Pneumothorax	35	31	88.57
Tuberculosis	31	29	93.55
	206	186	Average: 90.29%

To evaluate the proposed system of Lung Disease Recognition, a table of inputs data had been created. To each kind of lung disease, approximately 206 X-ray samples from testing sets were used to test. Table 1 shows the experimental results for the proposed system. The average recognition accuracy of proposed system is 90.29%.

In this study, the best result of the classification accuracy is in Pleural Effusion disease. The second best result for the classification accuracy was obtained in COPD X-ray samples. The classification accuracy performance on Tuberculosis was similar and closer to that of Pneumonia. From the same table, it can be seen also that the results obtained in Normal Xray is not that accurate, for the fact that it only gathered 79.31% of accuracy in the Neural Network System.

4. CONCLUSIONS AND RECOMMENDATIONS

The Neural Network is successfully implemented since the network showed fast and reliable results in identifying lung diseases. Out of 206 X-ray images, 186 of those were correctly categorized according to their particular disease classification. This is significant for it will reduce the long hours of patients waiting for their results and it also reduces the amount of load for medical experts. Although the network functions properly, an attending doctor will still be needed for the result verification.

In order for the neural network to process the data inputs properly several factors should be considered. The number of neurons and input affects the functionality of network. The default number of neurons or sample input to be processed is ten, thus decreasing the number of neurons may cause undesired effect to the network. The neural network's capability to recognize a particular lung disease depends on the type and number of lung disease fed on the neural network in the form of X-ray images. Since data to be inputted in the network will be in the form of images, these images must be clear enough to be identified as specific parts or organs and the details must be well-defined. The neural network is also sensitive to the quality of the image, thus a high-definition camera with a 10-MP (mega pixels) lens is recommended but a more defined camera lens will be more appropriate. Another consideration is the angle to which the X-ray was taken, a full overview of the lung must be properly presented. The background lighting of the X-ray film also affects the data presentation; therefore it must be well-lighted. If it is possible, avoid unnecessary image noise due to use of filter, etc.

The number of the data input in the form of an image per disease category must have more than twenty image samples for a precise result. Other kinds of disease related to the lungs may be added to the neural network to allow more functionality. Lastly, trial and testing after changing the number of neurons in the neural network is required in order to compare the previous outcomes and single out which is the best performance of the network. Then the network must always be trained a number of times until the maximum goal performance is reached.

5. REFERENCES

American Lung Association. 2008. Retrieved from Lung Disease Data: 2008: www.lungusa.org

World Health Organization. 2008. Retrieved from www.who.int

RadiologyInfo.org. 2014. Retrieved August 31, 2014, from Radiological Society of North America, Inc. (RSNA): http://www.radiologyinfo.org/en/info.cfm?pg=chestrad

Barúa, M., and H. Nazeran. 2004. Classification of Pulmonary Diseases Based on Impulse Oscillometric Measurements of Lung Function Using Neural Networks.

Gunasundari, S., and S. Baskar. 2009. Application of Artificial Neural Network in identification of.

Lakshmi, D. 2013. Comparison of Texture Analysis in the differentiation of Carcinoma from Other Lung Abnormalities Using Low-Dose CT Images.

Liang, T., and T. Toshiyuki. 2006. Automated Extraction and Diagnosis of Lung Emphysema from Lung CT.

Malathi, A., and A. Santra. 2012. Diagnosis of Lung Cancer Disease using Neuro-Fuzzy Logic.

Mayo Foundation for Medical Education and Research. (n.d.). Retrieved 2014, from Mayo Clinic: http://www.mayoclinic.org

Rubins, J. (n.d.). *Medscape*. Retrieved 2014, from Medscape: http://www.emedicine.medscape.com

Tadashi, K. and U. Junji. 2011. Medical Image Diagnosis of Lung Cancer by Revised GMDH-type Neural Network Using Heuristic Self-Organization.

Vinod, K. and S. Anil. 2013. Detection system for lung cancer based on neural network: X-Ray validation performance.