Israa S. Abed
Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Iraq.
eng.biomedical599@gmail.com

Received on: 24/09/2018
Accepted on: 24/01/2019
Published online : 25/05/2019

Lung Cancer Detection from X-ray images by combined Backpropagation Neural Network and PCA

Abstract-*The lungs are portion of a complex unit, enlarging and relaxing numerus times every day to supply oxygen and exude CO2. Lung disease might occur from troubles in any part of it. Carcinoma often called Cancer is the generally rising and it is the most harmful disease happened in humankind. Carcinoma occurs because of uncontrolled growth of malignant cells inside the tissues of the lungs. Earlier diagnosis of cancer can help save large numbers of lives, while any delay or fail in detection may cause additional serious problems leading to sudden fatal death. The objective of this study is to design an automated system with an ability to improve the detection process in order to perform advanced recognition of the disease. The diagnosis techniques include: X-rays, MRI, CT images etc. X-ray is the common and low-cost technique that is widely used and it is relatively available for everyone. Rather than new techniques like CT and MRI, X-ray is human dependable, meaning it needs a Doctor and X-ray specialist in order to determine lung cases, so developing a system which can enhance and aid in diagnosis, can help specialist to determine cases in easily.*

Keywords- *Lung Cancer, Intelligent Systems, Classification, Feature Extraction, Pattern Recognition.*

How to cite this article: I.S. Abed "Lung Cancer Detection from X-ray images by combined Backpropagation Neural Network and PCA," *Engineering and Technology Journal*, Vol. 37, Part A, No. 05, pp. 166-171, 2019.

1. Introduction

One of the main reasons for deaths both in women and men is the Lung Cancer. In many cases indication of lung cancer in a patient's body is divulged by early symptoms. Prognosis and the treatment depend on the kind of cancer, the cancer stage and the patient performance. The probable therapies involve surgery, radiotherapy and chemotherapy. Nowadays there are not any technical techniques to stop cancer; the only effective way is to detect cancer in early stages that is an acute factor in the therapy and improving the rate of survival [1].

Since lung cancer has become the common cancers in the world, the issues of being focused on quality of life and also survival need to be understood. Early detection is the main strategy for decreasing the death caused by lung cancer. Nowadays, three main methods have been used for lung cancer diagnosis which are: biochemical diagnosis, cytology histology diagnosis and imaging diagnosis, where the third one are the common and effective method for detection lung cancer [2].

The most favourite widely used techniques for diagnosing the lung cancer. Presently, various techniques are available for scanning the chest and diagnosing the lung cancer. With modern developments in X-ray, MRI, CT and ultrasonography (US). A choice of the ideal test may be difficult, and the selection is related to the doctor to determine the best method to make sure that the most effective technique is used.

The common and main method that is used by doctors for first diagnosis is the X-Ray due to its fast result and low cost, but it has the disadvantage of poor contrast and difficulty to diagnose different lung cases [3].

Image processing methods are generally utilized usually in several medical fields, which offer a high-quality tool for enhancing the manual analysis. Furthermore, artificial neural networks (ANN) offer various strategies to solve problems and is called the 6th generation of computing. The image processing has been commonly used in the area of medical imaging. For these tasks, image processing had been utilized with the assistance of artificial intelligence tools for example backpropagation neural network (BP NN) to achieve the highest and the most accurate results. ANN have an excellent performance in classification and performance approximation, and it used successfully in medical image processing in the last years, especially in the case of: segmentation, pre-processing (e.g. restoration and construction), and recognition. The BP presents the most areas in pattern recognition field. The few other neural techniques such as fuzzy, convolution, radial basis function, Adaptive resonance theory, Hopfield and Probabilistic also have found their particular

http://dx.doi.org/10.30684/etj.37.5A.3
2412-0758/University of Technology-Iraq, Baghdad, Iraq

position in medical image recognition and detection [4, 5].
The goal of this paper is to propose a system for detection lung cancer that based on ANN and PC along with their related analysis approaches.

2. Related Works

Polat et al. [6] have proposed a method for detected carcinoma of the lung by using artificial immune recognition system (AIRS), fuzzy weighting pre-processing and the principles component analysis (PCA). This system consists of three levels. First, a dataset of lung carcinoma that has Fifty-seven features had been minimized to four features by using PCA. Next, a pre-processing step via weighting scheme that is based on fuzzy weighting pre-processing has been used prior to main classifier. Finally, AIRS was utilized as a classifier. Experiments had been conducted for the dataset of lung carcinoma to diagnose it with a completely automatic manner. The classification accuracy obtained of system reached up to 100% and it proved very promising intended for additional classification applications.
Ms I. Christa Mary et al. (2014) [7] presented a detailed survey of early detection of lung diseases by used Computer Aided Diagnosis system (CAD). In their work, a computerized classification and detection of lung image has been used which consisted of five methodologies, which are: preprocessing stage, segmentation stage, feature selection stage, feature extraction stage and classification stage. The comparisons using these procedures have been described in their study.
Abood (2014) [8], study the improvement of a breast cancer screening system that can be used by cytologists to differentiate between benign and malignant types using images that are typical of those currently interpreted by cytologists world-wide. The approach is considered based on features vector which is composed of Euclidian geometric parameters such as the object perimeter, area and infill coefficient in segmented cells of optical images of breast. The aim of study to create a system for classification of breast cancer, which is used by professional cytology of separation between benign and malignant cases. Medical images were analyzed used a global scale and widespread. The method used in the study based on a number of factors such as Euclidean engineering parameters, diameter, space and filling factor for cells withheld images from the visual images of the breast, and then depending on the rating of backpropagation neural networks.

3. Proposed Image Processing System

The proposed lung tumour detection method has the following steps:
1. Acquisition of X-Ray Lung Images
2. Pre-processing of X-Ray Lung Images
3. Feature Extraction by Principle component analysis
4. Neural network classifier
5. Diagnosis result

I. Image Acquisition

Clinical X-Ray patient's chest images has been used to build trained database, which consisted of 196 images (100 of cancer case and 96 non-cancer "normal" case) in addition to 196 images put aside to uses for test the recognition accuracy for no-trained images. These images are real cancer and normal x-ray images for patients and they have been collected from diagnostic center (Baghdad Radiotherapy and Nuclear Medicine Center-Medical City). The digitized images are stored as jpg format with a resolution of 512 x 512 pixels.

II. Processing the Images

After acquisition of lung images, all of these images have been processed in order to enhance details and prepared for training part. It includes many procedures that include: grayscale and resizing.

1) Greyscale Process

In this part we transformed the original RGB X-ray images to greyscale format. The goal from this process is to eliminate the tint information and saturation while leaving the luminance.

2) Resizing

After converting lung images, it was then resized from its original resolution to small symmetric size of (64x64) and (128x128) pixels. This part is important to trained images with good information in lower processing time.

III. Feature Extraction

Principles component analysis is very suiTable for reducing the number of factors that is included in a dataset along with leaving the dissimilarity inside the data and to identify anonymous patterns from the data and also classify them based on the quantity of the details, saved in the data. The PCA makes it possible for finding out a linear alteration to reduce details from higher dimensional toward lower dimensional [9].

V. Classification by Neural Network

In this work, a standard, 3 layers backpropagation ANN is used as a recognition method for detection Lung Tumor. The input image is of two sizes, 64x64 pixels which has 4096 neurons and128x128 pixels that have 16384 neurons. The hidden neurons are 70 and the output labels are two which is to lung has tumor and lung has no tumor. The binary code of lung with tumor is put as [1 0] while the lung has no tumor put as [0 1]. The sigmoid function is used regards the neurons activating purpose for both hidden layer and output layer. The backpropagation ANN is shown in Figure 1.

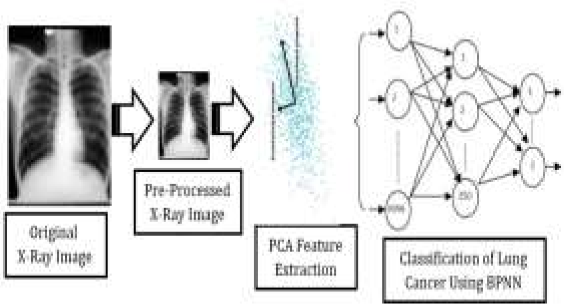


Figure 1: Lung tumor detection process

The algorithm of proposed system can be summarized as follows [10]:

1- Identify the network structures

o Identify the input neurons and output neurons (layers)

o Identify hidden neurons

2- Initialized the sigmoid function of the neural network. The quantities of threshold units must not change.

3- Chosen the pair of input and the output. Suppose the vector of input is X_i and assume the target output vector is Y_i.

4- Compute the sigmoid function from input units to the hidden units via employee the sigmoid equation as in follows [11]:

$$\Delta h_j = \frac{1}{1+e^{-\sum_{i=0}^{B} W1_{ij}}} \{\text{Forj=1…B}\} \quad (1)$$

Where, $W1_{ij}$ is the weight threshold for the vector j, and the i vector is range from 0 up to B.

5- Propagate the sigmoid function from the hidden units to output units via employing the sigmoid equation as in follows [9]:

$$\Delta h_j = \frac{1}{1+e^{-\sum_{i=0}^{B} W2_{ij}}}, \{\text{For j=1…,B}\} \quad (2)$$

Where, the $w2_{ij}$ is the weight threshold for the vector j, which is very important in the summation of weights.

6- Computes the errors of the units within the output layer denoted $\delta 2_j$. The errors of the units are dependent upon the actual output of the neural network (o_j) as well as, the target output (Yi):

$$\delta 2_j = o_j(1-o_j)(y-o_j), \text{For j = 1, …., B.} \quad (3)$$

7- Computes the errors of the units in hidden layer,

$$\delta 1_j = h_j(1-h_j)\sum_i^c \delta 2_i \times w_{ij}, \text{For j = 1, …., B, i= 1, ….} \quad (4)$$

$$\delta w1_{ij} = \eta \times \delta 2_j \times h_j, \text{For all i = 0, …., A, j=1…., B} \quad (5)$$

8- Adjust the weights amongst hidden layer and output layer. The learns rate signify the denoted by η; and its functions are identical to perception learning. A reliable value of f η is 0.35:

$$\delta w2_{ij} = \eta \times \delta 2_j \times h_j, \text{For all i = 0, …., B, j=1….,C} \quad (6)$$

9- Adjust the weights amongst input layer and hidden layer.

$$\delta w1_{ij} = \eta \times \delta 1_j \times h_i, \text{For all i = 0, …., A, j=1….,B} \quad (7)$$

10-Moving back to step 3 and perform steps once again. Once all of the inputs pairs and output pairs are totally provided to the network, a 1 epoch has been achieved. Likewise, reiterating steps right from 3 to 10 considering the total number of epochs is needed.

The method for this suggested procedure can be described as follows:

1. Gathering up several lung cancer real images to generate a database of lung images.
2. The image processing methodology is made up of transforming the RGB images to the grayscale images and then resizing it to preferred scale that will make the extra process easy.
3. PCA can then employed for the extraction of feature from the images. These features are going to enhanced by utilizing Bacterial Forging Optimization.
4. ANN is trained by making use of the extracted data from the images. The case of the lung (to be has tumor or not has) is classified by utilizing ANN.
5. The efficiency factors which include: False Acceptance Ratio, False Rejection Ratio and Accuracy are evaluated.

4. Results and Discussion

In this part, we have used the two input images of 64x64 and 128x128. We have trained them both, then compared with traditional backpropagation method in order to investigate the gain in both time and recognition accuracy. For this test we have used a high computational laptop that has 6 core CPU type (intel core i7 8750) of 4GH and 16GB Ram, 256 SSD HDD, and Nvidia GTX 1060 6GB graphics card at First, we run the PCA to visualize the X-rays and got the dimension reduction. The result of employing reduction by PCA is illustrated in Figure 2.

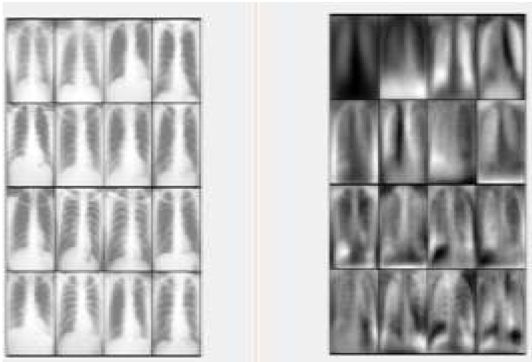


Figure 2: Image Processing by PCA: (a) original X-Ray lung images, (b) The X-Ray lung images after applying Dimension Reduction by using PCA

The PCA is worked to reduce the input neurons of original images, where the results show that the PCA has been reduced the input neurons for 64x64 images from 4096 neurons to 100 neurons as shown in Figure 3.

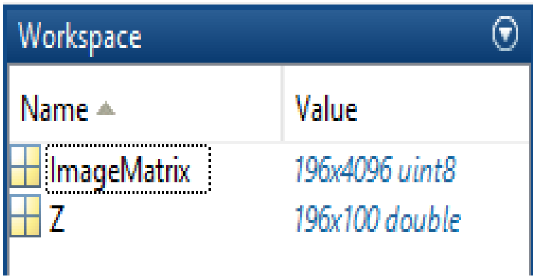
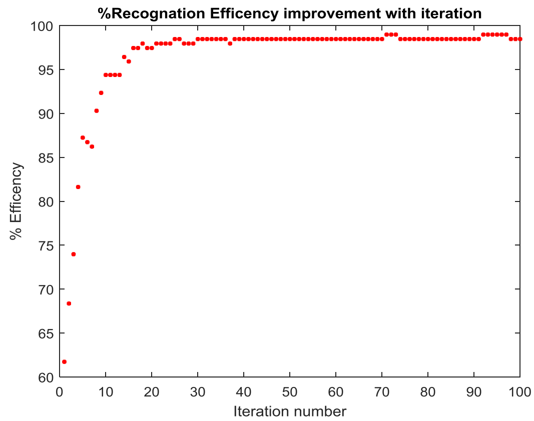
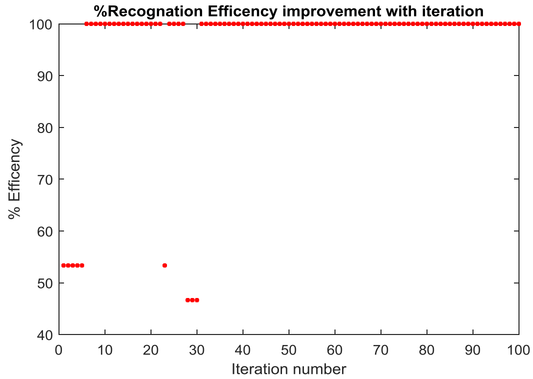


Figure 3: The input neurons for 196 images of 64x64-pixel size. Where, Image Matrix, the original input neurons for image, Z reduction of input neurons after applied PCA method

Then, we had tested the features extracted from PCA by using backpropagation ANN, the random initial weights are taken in range of (-1 to 1). To be able to obtain wanted minimal value of errors, we had chosen the training iteration to be 100. The results of training ANN for both 64x64 and 128x128 image sizes are shown in Figure 4, which represented the improvement of recognition (error reduction) per iteration.



(a)



(b)

Figure 4: The efficiency of ANN Recognition with iteration for image sizes (a) 64x64, (b) 128x128. The x-axis represents the efficiency while the y-axis represents the number of iterations

From training results shown in Figure 4, the recognition of trained images increases with increment of iteration that is related to decrease of cost function. The recognition is then reached up to 100% and it has been sTable after 32 iterations. The higher accuracy got is related to utilize of PCA. Figure 5, illustrated the recognition result of 16 sampled of lung x-ray images.

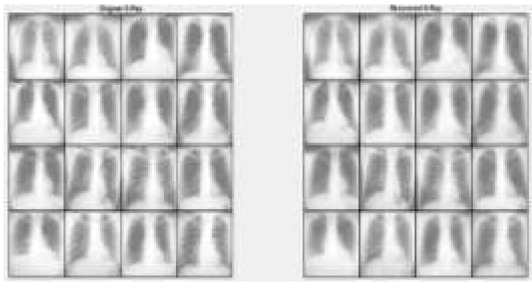


Figure 5: Recognition results where (left) is the original images and (right) the recognized images

From results shown in Figure 5, it appears that the original trained images (in left) are identical to recognized images (on the right), that means it achieved 100% accuracy percent in detection of trained images. To verifying the real results of recognition we have used non-trained images that determine its case by specialist in order to check the accuracy of detection. The recognition result appears that the proposed system achieved higher accuracy of detection lung cases (has or has no tumor). Figure 6 shows the diagnosis result for non-trained x-ray images and Table 1 shows the recognition results of trained and untrained lung X-ray images.

Table 1: Recognition Results for trained X-Ray images

Case	Size	Detect	Failed	Eff.
Cancer	64x64	97	3	97%
Cancer	128x128	100	0	100%
Normal	64x64	91	5	94.8%
Normal	128x128	94	2	97.92%

Figure 6: Recognition result for untrained X-Ray lung image, which has tumor

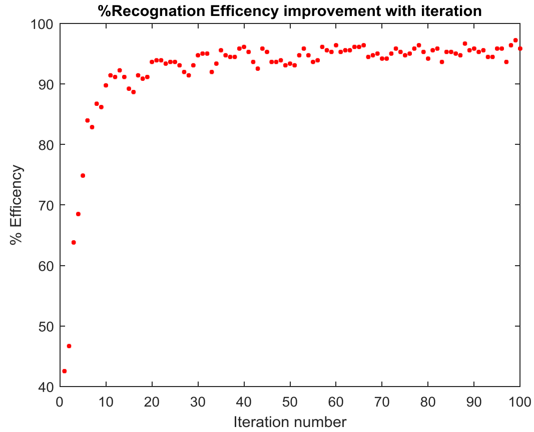
Table 2: Recognition results for non-trained X-Ray images

Case	Size	Detect	Failed	Eff.
Cancer	64x64	89	11	89%
Cancer	128x128	96	4	96%
Normal	64x64	82	14	85.42%
Normal	128x128	87	9	90.63%

For compaction with slandered BPNN, we have been tested both methods in term of time requested for training the network and the efficiency of NN. The results have shown in Table 3 and Figures 7 and 8.

Table 3: Training Time and efficiency results for BPANN system and the PCA-BP system

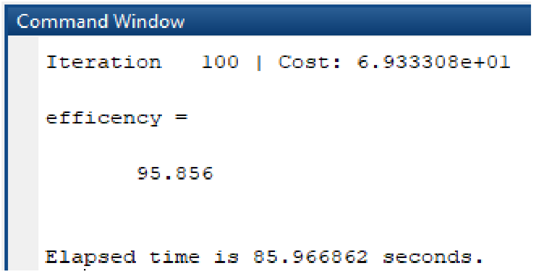
BPNN				PCA-BP			
64x64		128x128		64x64		128x128	
Time	Eff.	Time	Eff.	Time	Eff.	Time	Eff.
86 Sec	95.85%	159 Sec	97.3%	8 Sec	98.5%	17 Sec	100 %



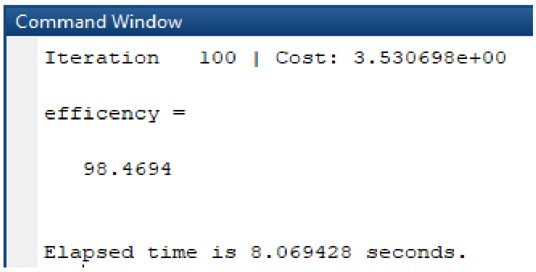
(a)

(b)

Figure 7: Training efficiency results when trained 64x64 images by used(a) BPNN (b) PCA-BP



(a)



(b)

Figure 8: Training time and efficiency result for system used (a) BPNN (b) PCA-BP

As shown from Table 3 and Figures 7 and 8, the PCA results are more effective than slandered in both times required for training network and the training accuracy, where the result for 64x64 images shows that PC-BP got a 98.5% efficiency in very fast time of 8 second which is ~3% higher efficiency and 76 sec faster than slandered BP. For 128x128 images, the results show that PC-BP got a 100% efficiency in very fast time of ~16 second which is ~3% higher efficiency and ~142 sec faster than slandered BP. This result proved that the using of PCA can improve the recognition of ANN in both times requested and accuracy as well as the computational requirement for these processes even with high input nodes because it reduces the input neurons and as a result the computational requirement.

VII. Conclusion

In this work, we have proposed detection system that can detect the lung tumor by utilized (PCA) for feature extraction with traditional backpropagation neural network (BPNN) and has been presented. The system is used for detection the lung situation if it has tumor or without not. The preprocessing of images is aimed to give significant representations of lung patterns while minimizing the amount of data, therefore minimizing both the computational costs and time required. The main benefit from utilized PCA is to minimize the dimension of trained images and improving the recognition results of ANN along with reduced the execution time. With regards of uses untrained images, the overall accurate recognition results of 95% has become achieved, in total amount of 95 images out of 100 for x-ray images with tumor and 96.67% has been obtained, in total amount of 87 images out of 90 for x-ray images without tumor, have been accurately identified. This proposed system can be utilizing to assists radiologists to easy classify the lung case. The results approved that uses of (PCA) with BP ANN is better than uses just BP ANN alone in detection of lung tumor.

Reference

[1] T. Kaur and N. Gupta, “Classifications of Lung Diseases by uses Optimization Techniques,” International Journal for Scientific Research and Development (IJSRD), Vol.3, Issue.8, 2015.

[2] K. Balachandran, and R. Anitha, “An Efficients Optimization Based upon Lung Cancer Prediagnosis System with Aided of Feed Forward Back Propagation Neural Network (FFBNN),” Journal of Theoretical and Applied Information Technology, Vol.56 No.2, 2013.

[3] G.W. Staton and S. Bhalla, “Imaging of Lung Disease,” BC Decker Inc., DOI 10.2310/7900.S14C02, 14 resp ii, 2008.

[4] K. Dimililer, Y.K. Ever and B. Ugur, “ILTDS: Intelligent Lung Tumor Detection System on CT Images,” Springer International Publishing, Intelligent Systems Technologies and Applications, DOI 10.1007/978-3-319-47952-1_17, 2016.

[5] U.I. Dike and U.A. Adoghe, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential,” International Journal of Computers and Distributed Systems, Vol. No.3, Issue 2, ISSN: 2278-5183, Jun-July 2013.

[6] P. Kemal and G. Salih, “Principles component analysis, fuzzy weighting preprocessing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer,” Expert Systems with Applications, Vol.34, No. 1, pp. 214–221, 2008.

[7] Z.M. Abood, “Benign and Malignant of Breast Tumors Classification by Backpropagation Neural networks,” Iraqi Journal of Information Technology, Vol. 6, Issue 2, pp.13-20, 2014.

[8] I.C. Mary, J. Preethi, “A Survey on Computerized, Quantification and Classification of Lung Disease,” IJAICT, 2014.

[9] U. Javed, M. M. Riaz, T.A. Cheema and H.M. Zafar, “Detection of Lung Tumor in CE CT Images by using Weighted Support Vector Machines,” IEEE, pp. 113-116, 2013.

[10] S.K. Devireddy and S.A. Rao, “Hand Written Character Recognition Using Backpropagation Network,” Journal of Theoretical and Applied Information Technology (JATIT), 2009.