Vol. 2 No: 1 3rd year 2018 Al-Kut Univ. College Journal مجلة كلية الكوت الجامعة
ISSN: 2414 – 7419

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

Israa Safaa AL-Zubaidi
Al_Khwarizmi College of Engineering, University of Baghdad
eng.biomedical599@gmail.com

Abstract

The lungs are portion of complex unit, enlarging and relaxing many, many times every day to supply oxygen and exude CO_2. Lung disease might occur from troubles in any part of it. Carcinoma often called Cancer is the generally rising and it the most harmful disease happened in human kinds. Carcinoma occurs because of uncontrolled growth of malignant cells inside the tissues of the lungs. Earlier diagnosis of the cancer can helps 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 steady is to design a automated system have ability to improve the detection process in order to perform advanced recognition of the disease. The diagnosis techniques includes: X-rays, MRI, CT images etc. X-ray are the common and low cost technique that are widely used and it relatively available for everyone. Rather than new techniques like CT and MRI, X-ray it human dependable, which mean it need to Doctor and X-ray specialist in order to determine lung cases, so developing a system can enhance and aided in diagnosis can helping specialist to determine cases in easy.

In this proposed system, we have been used a technique that include Image Preprocessing, Training by neural network and recognition of samples which are essential for the task of medical image mining. A highly effective training model i.e. (BPNN) has been used for classification, which could classify digital X-ray, MRI, CT images, etc. to normal or cancerous lung. Additionally, to enhancing training process, PCA method has been utilized to improve tumor detection. This technique may give a good diagnosis about a patient's state that can be utilized in medical imaging application.

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

كشف سرطان الرئة من الأشعة السينية من خلال استخدام طريقة الـ (Backpropagation Neural Network) و الـ (PCA)

أسراء صفاء الزبيدي
كلية الهندسة الخوارزمي / جامعة بغداد
البريد الألكتروني: eng.biomedical599@gmail.com

الخلاصة

الرئة هي وحدة معقدة، تتسع وتتقلص مرات عديدة كل يوم لتجهيز الأوكسجين والتخلص من ثنائي اوكسيد الكاربون. أنامراض الرئة ممكن ان تحدث نتيجة مشاكل في اي جزء من أجزائها. يعد السرطان (Cancer) من الأمراض الواسعة الأنشاروالأكثر خطورة التي تواجه الجنس البشري، ويحدث بسبب النمو غير المسيطر للخلايا الخبيثة داخل نسيج الرئة. التشخيص المبكر للسرطان يمكن ان يساعد على انقاذ ارواح عديدة من الناس، حيث ان اي تاخير او الفشل في كشفه يمكن ان يسبب المزيد من المشاكل المتسلسلة التي تؤدي الى الموت المفاجئ. ان هذه الدراسة تهدف الى تصميم نظام اوتماتيكي له القابلية على تحسين طرق الكشف لكي يوفر طريقة تمييز متقدمة لهذا المرض. طرق التشخيص تتضمن: الأشعة السينية (X-rays)، الرنين المغناطيسي(MRI)، والأشعة المقطعية (CT)، وغيرها. يعد التشخيص بلأشعة السينية هو من اتقنيات الشائعة الأستخدام والرخيصة الثمن التي تستخدم بصورة واسعة ومتاحة لكل شخص. على النقيض من التصوير بلرنين المغناطيسي والأشعة المقطعية فأن الأشعة السينية تعتمد على الأنسان، اي انها تحتاج طبيب ومتخصص بلأشعة السينية لتحديد حالة الرئة المرضية، لذا فأن تطوير نظام يحسن ويساعد في التشخيص يمكن ان يساعد المتخصصين في تحديد الحالة المرضية بسهولة.

في هذا النظام المقترح، نستخدم تقنية تتضمن المعالجة الصورية, التدريب بالشبكة العصبية ومن ثم التمييز والالتي تعتبر اساسية في مجال التصوير الطبي. تم استخدام موديل تدريبي عالي الكفاءة (BPNN) للتصنيف والذي يسمح بتصنيف الصور الطبية (السينية أو الرنين المغناطيسي او الاشعة المقطعية وغيرها.) الى سليمة او سرطانية. أضافة الى تحسين عملية التدريب، تم استخدام طريقة (PCA) لتحسين كشف الورم السرطاني. هذه الطريقة يمكن ان تعطي تشخيص جيد عن حالة المريض والتي يمكن ان تستخدم في التطبيقات الطبية التصويرية.

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 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 is the common and effective method for detection lung cancer [2].

The most favourite and widely used techniques for diagnosing the lung cancer are available for scanning the chest and diagnosis the lung cancer with modern developments in X-ray, MR, CT and ultrasonography (US). A choice of the ideal test may be difficult, and the selection is related to a doctor in order to determine the best method to make sure that the most effective technique is used. The common and main method that used by doctors for first diagnosis is the X-Ray due its fast result and low cost, but it has disadvantage of poor contrast and difficult to diagnosis 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 very various strategy to solving problems and 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 a few other neural techniques such as fuzzy, convolution, radial basis function, Adaptive resonance theory, Hopfield and Probabilistic also have found their particular position in medical image recognition and detection [4, 5].

In this research we have been designed intelligent lung tumour detection system based on backpropagation neural net and Principal Component Analysis (PCA). 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 used PCA. Next, a pre-processing step via weighting scheme that 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 was reach to 100% and it approved it been a 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 pre-processing stage, segmentation stage, feature selection stage, feature extraction stage and classification stage. The comparisons using these procedures have been described in there steady.

3. Proposed Image Processing System

The proposed lung tumour detection method has 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

ISSN: 2414 – 7419

3.1 Image Acquisition

Clinical X-Ray patient’s chest x-ray images has been used to build trained database, which consisted of 196 images (100 of cancer case and 96 non-cancer case) that have been collected from diagnostic center. The digitized images are stored as jpg format with a resolution of 512 x 512 pixels.

3.2 Pre-processing of Images

After acquisition of lung image, each of these images have gone little pre-processing procedures including normalization, grayscale conversion, etc.

3.2.1 Grayscale Conversion

This method requires conversion of original X-ray images that are in RGB format to grayscale by eliminating saturation and tint information while keeping the luminance.

3.2.2 Normalization

After grayscale an X-Ray Lung Images, it then converted from its original size to (64x64) and (128x128) pixels’ values. This specific size provides reliable information in low processing time.

3.3 Feature Extraction

PCA is generally ideal for lowering the number of factors that comprises a dataset whilst keeping the inconsistency within the data and to recognize unknown patterns from the data as well as categorize them based on the amount of the information, saved in the data. PCA enables determining a linear alteration of which maps information by a higher dimensional space into a lower dimensional space [8].

3.4 Neural Network Classifier

A standard, 3-layer BPNN has been utilized in Intelligent System for Detection Lung Tumour with 64x64 and 128x128 input images, that have 4096 and 16384 input neurons respectively, and used 70 hidden neurons and two output neurons that categorise the lung to lung with tumours and with no tumours. The lung is sorted in binary coding as [1 0] lung with Tumour and [0 1] for the lung without Tumour. The activation function (sigmoid) has been used for activating purpose of the neurons both in output and hidden layers. The BP network is illustrated in figure 1. Due to the simplicity of implementation, as well as the availability of adequate database (input target") for training, utilizing a BPNN that is a supervised learner, is often preferred. Generalization (and Training testing are made in this stage.

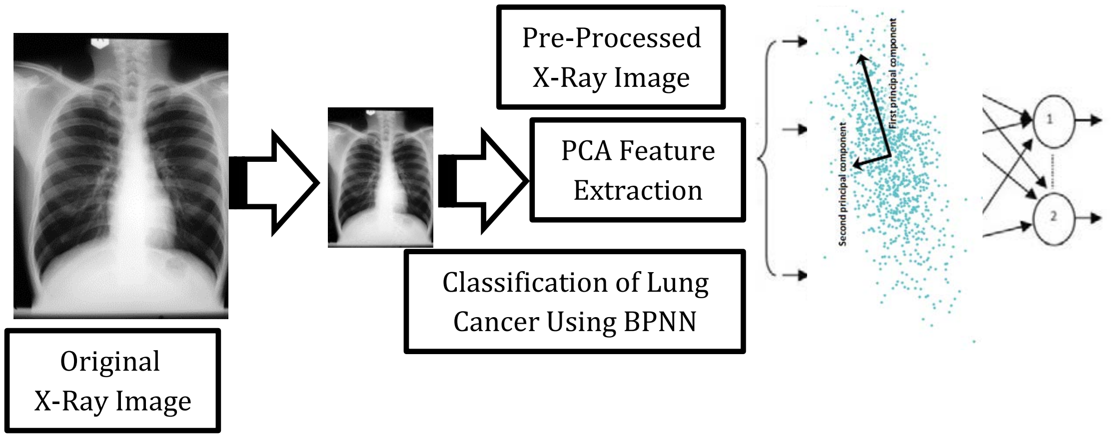

Figure 1: Lung tumour detection process.

The backpropagation algorithm can be summarized as follows [9]:

Step 1: Determine the network architecture

- Determine the input and output neurons; and output labels
- Determine hidden neurons and layers

Step 2: Initialized the activation of the neural network. The thresholding unit's values should do not change. a) X = 1.0 b) h = 1.0

Step 3: Select an input and output pair. Assume the input vector is Xi and assume the Yi is the target output vector. Give activation levels for the input units.

Step 4: Propagate the activation function from input units to the hidden units by utilizing the activation functions [9]:

$$\Delta h_j = \frac{1}{1+e^{-\sum_{i=0}^{B} W1_{ij}}} \quad \text{For } j=1, \ldots, B \qquad (1)$$

The i ranges from 0 to B. $W1_{ij}$ is the thresholding weight for j.

Step 5: Propagate the activation function from the input units to the hidden units using the activation functions [9]:

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

The thresholding weight $w2_{ij}$for output unit j is important in the weighted summation. where h is 1.

Step 6: Calculate the units' errors inside the output layer denoted $\delta 2_j$. The Error is dependent on the network real output (o_j) also, the target output (Yi):

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

Step 7: Calculate the units errors in the hidden layer, denoted $\delta 1_j$

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

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

Step 8: Modify the weights between hidden and output layer. The learning rate denoted represent the denoted by η; and it is functions are similar to perception learning. A sensible value of η is 0.35.

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

Step 9: Modify the weights between input and hidden layer.

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

Step 10: Return to step 3 and do steps again. As soon as all of the inputs/output pairs have been completely given to the network, a one epoch has become accomplished. Also, repeating steps from 3 to 10 based on how many epochs is desired.

The procedure for this proposed procedure can be summarized as follows:

1. **Step 1:** collecting various lung cancer images to make a database of these images.

2. **Step 2:** The pre-processing method, which consists of converting images to grayscale and normalizing it to desired scale, which makes the further proceedings easy.

3. **Step 3:** Principal Component Analysis will then be applied for the feature extraction of the images. Features will be optimized by using Bacterial Forging Optimization.

4. **Step 4 :** The neural network is trained by using the data extracted from the images. The stages of the lung cancer will be classified by using neural network.

5. **Step 5 :** Performance parameters such as: Accuracy, False Rejection Ratio and False Acceptance Ratio has been evaluated.

2. Results

For first we run PCA and visualize the eigenvectors which are in this case Eigen-Rays, then visualization of X-Ray after PCA Dimension Reduction. The result of applying PCA reduction is shown in figure 2.

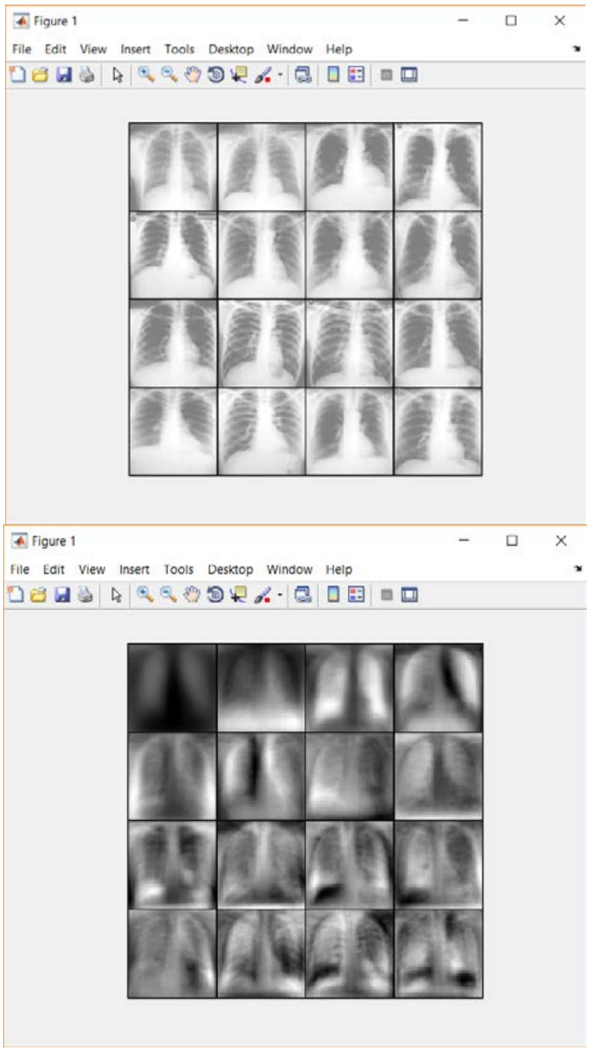

Figure 2: Processing Data: (a) original images, (b) Visualization of X-Ray after PCA Dimension Reduction.

We have bbeen trained the PCA data by used BPNN learning phase used initial random weights that has a values in range between -1 and 1. So as to achieve requested minimal error value we selected iteration to be 1000. The learning result has been shown in figure 3, which represented the improvement of recognition (error reduction) per iteration.

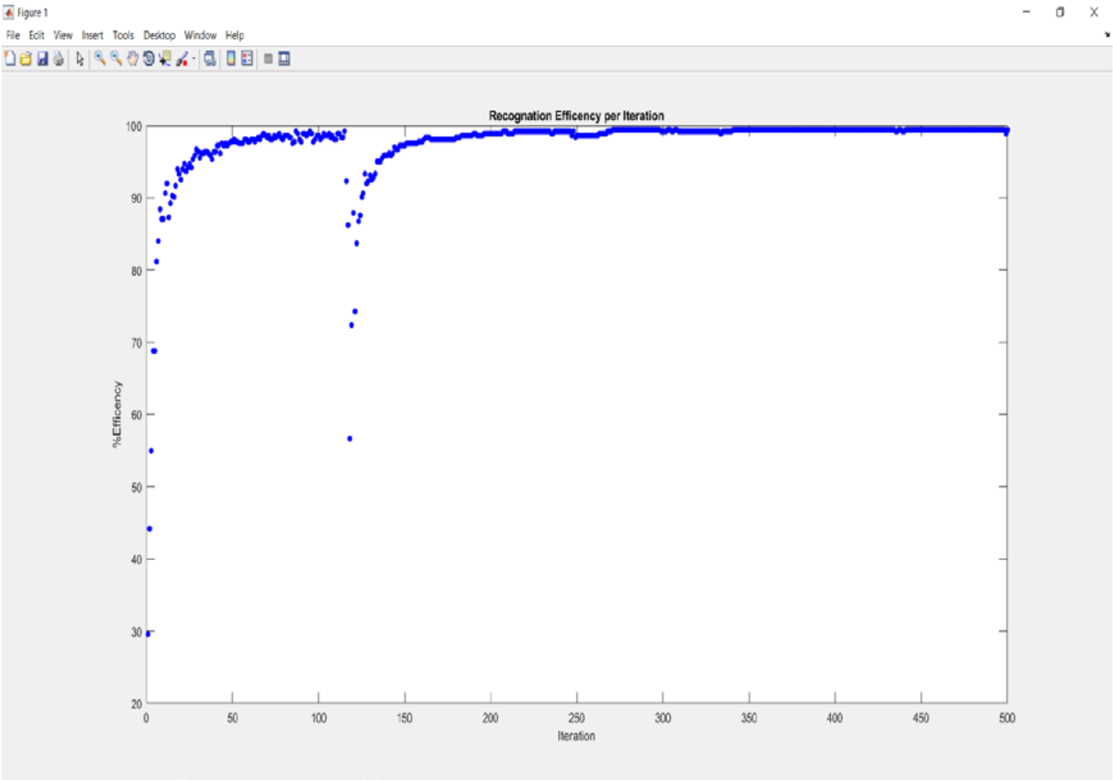

Figure 3: Recognition efficiency per iteration, x-axis the efficiency and y-axis the iteration.

As shown from figure 3, the efficiency of recognition has been increased with iteration as a result to reduction in cost function and it has been reach to stability after 350 iterations to be ~100%. this higher value proves the improvement in training process by PCA. Figure 4 shows side by side recognition result of sampled of x-ray images.

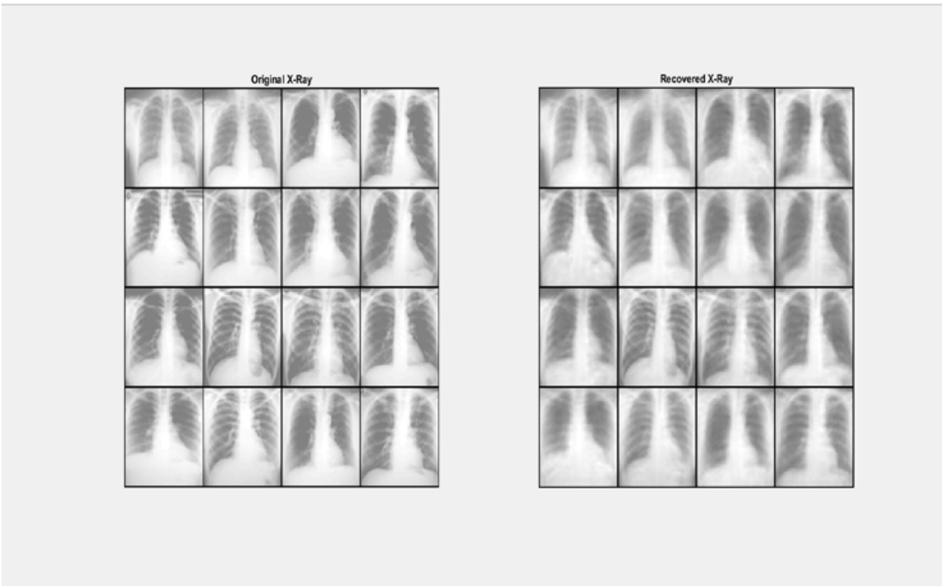

Figure 4: Comparison between to the original x-ray images (left) and the recognized image (right).

As shown in figure 4, where the image in left are original trained images and the left recognized output image that approved the recognition has a 100% detection.

To verifying the real recognition test we test the recognition for non-trained images that has been checked by special radiologist to determine the lung case. The result shows a high accuracy in recognition and detection lung that have tumors or normal one. Figure 5 shows the diagnosis result for non-trained x-ray images and table 1 shows the recognition results of trained and non-trained x-ray images.

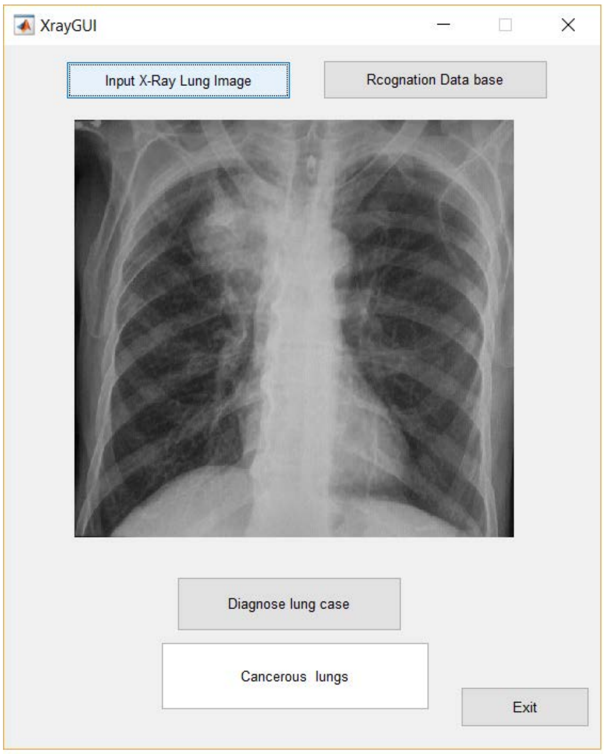

Figure 5: Recognition result for non-trained X-Ray image that have a tumor.

Table 1: Recognition Results for trained X-Ray images

Method		Image Normalization	X-Ray lung image Cases			
			Cancer (100)		Normal (90)	
			Detect	Failed	Detect	Failed
1	Trained	64 x 64	97	3	88	2
2	Non trained	64 x 64	89	11	76	14
3	Trained	128 x 128	100	0	100	0
4	Non trained	128 x 128	95	4	87	3

As shown from table 1, the trained images have high accuracy recognition for trained image and non-trained images, also the recognition error will reduction as the image matrix increases that due to increasing in details that improving learning results.

7. Conclusion

In this paper, a lung tumor detection system using (PCA) as a feature extraction method and backpropagation neural network (BPNN) and has been presented. The system has been utilized to diagnose the lungs with or without tumors. The preprocessing of Images is aimed to giving significant representations of lung patterns while minimizing the amount of data, therefore minimizing both the computational costs and time required. PCA has been used to reduce the dimension as well as improved the BP-ANN performance in terms of execution time. Successful system implementation, has been implemented, to identify the normal lung and lung that have tumors. Considering non unknown cased that represented by non-trained images, an 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 assists the doctors and radiologists for easy distinction between normal and cancerous lung. From our study, it provides that the using (PCA) with BPNN is more accurate than using it alone in detection of lung cancer.

References

1- Kaur T. and Gupta N., “Classification of Lung Diseases using Optimization Techniques”, International Journal for Scientific Research & Development (IJSRD), Vol. 3, Issue 08, ISSN: 2321-0613, 2015.

2- Balachandran K., Anitha R., “An Efficient Optimization Based Lung Cancer Pre-Diagnosis System with Aid of Feed Forward Back Propagation Neural Network (FFBNN)”, Journal of Theoretical and Applied Information Technology, Vol. 56 No.2, ISSN:1992-8645, 2013.

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

4- Dimililer K., Ever Y. K. and Ugur B., "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- Dike U. I., Adoghe U. A., "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- Kemal P. and Salih G., "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- Mary I. C., Preethi J., "A Survey on Computerized, Quantification and Classification of Lung Disease", IJAICT, 2014.

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

9- Devireddy S. K. and Rao S. A., "Hand Written Character Recognition Using Backpropagation Network", Journal of Theoretical and Applied Information Technology (JATIT), 2009.