Lung Cancer Classification using Curvelet Transform and Neural Network with Radial Basis Function

M. Obayya
Electronics and Communication
Engineering Department
Faculty of Engineering
Mansoura University
Mansoura, Egypt

Mohamed Ghandour
Electronics and Communication
Engineering Department
Faculty of Engineering
Mansoura University
Mansoura, Egypt

ABSTRACT

In this paper, computed tomographic (CT) chest images were investigated to develop an automated system to discriminate lung cancer. These were done by analyzing Data recorded for patients with benign cancer, and also patients with malignant lung cancer were taken in account. The techniques for utilized feature extraction included features derived from texture analysis based on Gray Level Co-occurrence Matrix (GLCM) of the input image, as well as features derived from curvelet transform-based features. An artificial neural network (ANN) with radial basis function classifier was utilized to classify the type of cancer whether benign or malignant. The results have shown that using curvelet domain features gives the highest rate to recognize lung cancer. Classification correct rate is up to 96%.

Keywords

Curvelet transform, lung cancer, image processing, ANN with radial basis function.

1. INTRODUCTION

Lung cancer is the independent growth of aberrant cells in one or both lungs. It is the number one cause of deaths in humans. Survival from lung cancer is proportional to its detection time. If it is detected in early stages it increases the rate of survival. According to the latest statistics, a total of 1,660,290 new cancer cases and 580,350 deaths from cancer are projected to occur in the US in 2013. Cancer that starts in the lung is known as primary cancer. The main types of lung cancer are small-cell lung carcinoma (SCLC), also called oat cell cancer, and non-small-cell lung carcinoma (NSCLC). The most common symptoms are coughing (including coughing up blood), weight loss and shortness of breath. Chest Radiograph (x-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology are the techniques to diagnose the lung cancer. In terms of speed, analysis of large volume of data CT images has more advantage over the other techniques [1].

Computerized tomography (CT) is the most frequently used imaging technique in the diagnosis of lung cancer. Nodules and pathological residues with varied diameter can be comfortably viewed by CT [2]. Nodules on the lung are classified as benign or malignant. During diagnosis, malignant nodules that are solid and atypical can be assessed as benign in some cases. However, in most cases, a solid nodule is usually classified as malignant [3]. It is crucial to diagnose nodules at early stages in order to accelerate the treatment process.

In the literature, there are some studies regarding early diagnosis of lung cancer and identification of nodules. E.Dandil et al. [4] developed a CAD system to provide segmentation of nodules on the lobes with neural networks model of Self-Organizing Maps (SOM) and ensures classification between benign and malignant nodules with the help of ANN using GLCM features. Performance values of 90.63% accuracy and 92.30% sensitivity were acquired in the CAD system which utilized a total of 128 CT images obtained from 47 patients. Yongjun et al. [5] reported method to classify the nodules based on 5 clinical parameters, using ANN and Logistic regression models; the accuracy was found out to be 84.6%, the results were compared by ROC curve analysis. J. Kuruvilla et al. [6] selected a computer-aided classification method in CT images of lungs by feed forward back propagation neural networks and statistical features. This work resulted in 93.3%, specificity of 100% and sensitivity of 91.4%. Amal Farag et al. [7] examined the effectiveness of geometric feature descriptors for false positive reduction and for classification of lung nodules in low dose CT (LDCT) scans using active appearance models (AAM); which are then used to detect candidate nodules based on optimum similarity measured by the normalized cross -correlation (NCC). R. C. Hardie et al. [8] combined Weighted multi-scale convergence index (WMCI) and fisher linear discriminant (FLD) and the system was evaluated on 154 chest radiographs from JRST database with 100 malignant and 54 benign nodules which results in the correct classification of 78.1% of the nodules in the JRST test set and average of 4.0 false positives per image.

The rest of this paper is organized as follows; Section 2 data collection. Section 3 briefly explains the methodologies adopted. Section 4 presents the experimental results and Section 5 gives the concluding remarks.

2. DATA COLLECTION

Performance of the proposed CAD system in the present study is evaluated using the LIDC database [9]. It is publicly available in the National Biomedical Imaging Archive (NBIA), and its nodules have been fully annotated by multiple radiologists. A total of 100 CT images of 512×512×8 bit from 50 different patients were collected. Based on the annotation Data, 50 of these nodules were benign and 50 were malignant. All images are available in Digital Images and Communication in Medicine (DICOM) format.

3. METHODS

The proposed system that processes and classifies automatically the digital CT images of the human lung consists of four main parts as shown in Fig.1. The first part is image enhancement to increase the quality of the input image. The second part is image segmentation to extract the lung region and ROI. The third part is feature extraction and selection to extract the main features of the tumor using different techniques in spatial and transformed domain. The last part is the classifier to discriminate the tumor whether it is benign or malignant.

3.1. Image Enhancement

Medical images are corrupted with noise and artifacts due to body movements. Thus wiener filter [10] of mask size 3*3 is used to remove noise as shown in Fig.2.

3.2 Segmentation of Lung Region and ROI

Segmentation is the process of separating the lung lobes from CT image and it classifies the suspicious regions. Accurate lung segmentation allows for the detection and quantification of abnormalities within the lungs. Segmentation in this paper is based on threshold and morphological operations [11]. The filtered CT image shown in Fig.3.a is converted to binary image using adaptive threshold and the result is shown in Fig.3.b. The inverse of the resulting image is taken then the borders are cleared as shown in Fig 3.c.and Fig. 3.d. The lung regions (left lung and right lung) are extracted by selecting the largest two areas as shown in Fig.3.e, Applying fill operation then multiplying the resulting image by the input image to get the lung as shown in Fig.3.f. The segmented lung image is then converted to a threshold image then a closure operation is applied to remove the small nodules and detect the tumor and the result is shown in Fig.3.g, Fig.3.h.

3.3 Feature Extraction and Selection

The third step is extracting the most salient features of tumor and reducing the dimensionality of acquired data [12]. Two sets of features were utilized and their performance was evaluated. The techniques used to calculate Features were: (1) texture features that refer to the characterization of tumor regions in an image by their texture content. These are entropy of grayscale image, contrast, correlation, energy, and homogeneity of the create gray-level co-occurrence matrix from image at $0^0, 45,^0 90^0$ and 135^0 angle directions in d=2 distance then the 4 results are averaged [13]. (2) Features derived from curvelet transform [14]. Fig.4 shows the FDCT coefficients derived using FDCT via wrapping. The low frequency coefficients are stored at the centre of the display. The Cartesian concentric corona shows the coefficients at different scales; the outer coronae correspond to higher frequencies. There are four strips associated to each corona, corresponding to the four cardinal points; these are further subdivided in to 16 angular panels. Each panel represents coefficients at a specified scale and orientation. In this study, the tumor image is decomposed into three levels or scales using real-valued Curvelets. The number of second coarsest level angles used is 16. This results in 1 subband at finest level (L=1), 16 subbands at second coarsest level (L=2) and another 1 subband corresponding to last coarsest level (L=3).
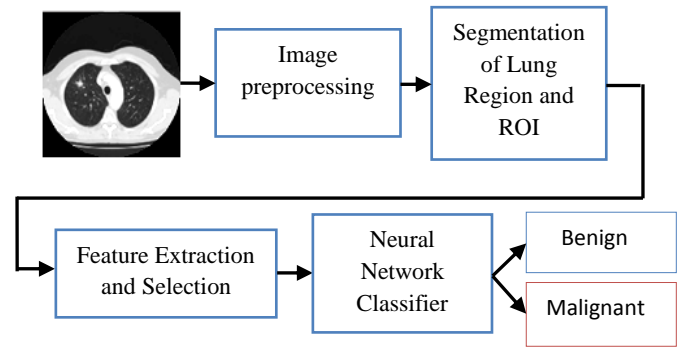

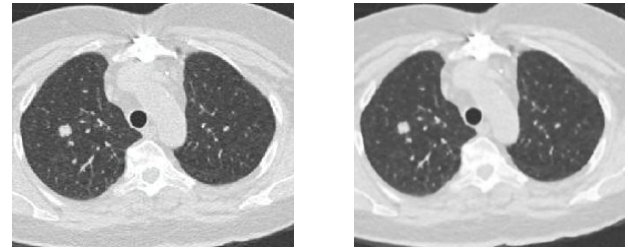
Fig.1 Block diagram of the proposed system

Fig.2: Left: Input CT Image, Right: The Filtered Image

The features like mean, standard deviation, energy and entropy for the obtained from each subband, features likes mean, variance, entropy and Energy are calculated. This will lead to a features vector of size $(4+4\times16+4) \times 1 = 72 \times 1$. This feature vector is fed as the input to the classifier for the purpose of Classification.
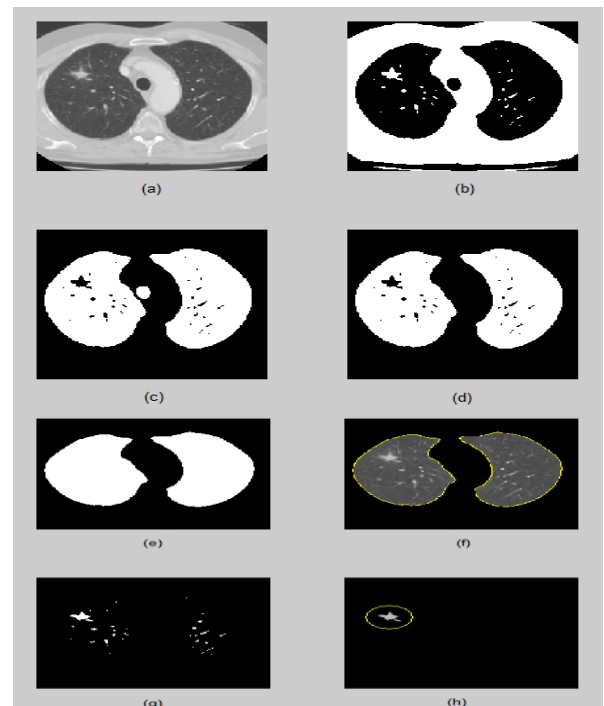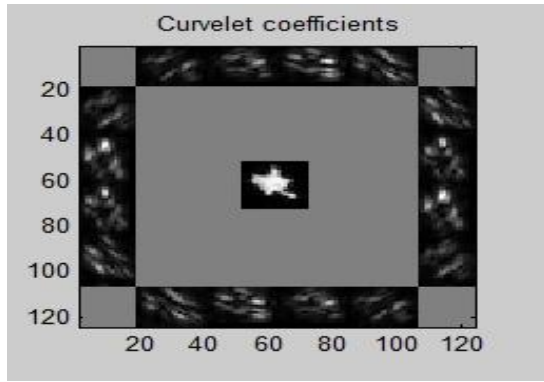

Fig.3: The steps of Lungs and tumor segmentation

Fig. 4 : FDCT Coefficients

3.4 Classifier Using Artificial Neural Network

The fourth step in the suggested system is to design an artificial neural network (ANN) classifier with radial basis function [15] which discriminate correctly the different types of human tumor lung. The input patterns to the classification stage are the resulting features vectors from the features extraction and selection stage. The classifiers were trained and their performance was evaluated using the dataset of CT lung images.

3.4.1. Network Topology

This network consists of three layers, input layer, hidden layer, output layer as shown in Fig.5. Input layer is composed of input nodes that are equal to the dimension of the input vector x. the network output can be obtained by

$$y = f(x) = \sum_{i=1}^{k} w_i \, \varphi_i(x) \qquad (1)$$

Where $f(x)$ is the final output, φ_i denotes the radial basis function of the *i-th* hidden node, w_i denotes the hidden-to-output weight corresponding to the *i-th* hidden node, and k is the total number of hidden nodes. A radial basis function is a multidimensional function that describes the distance between a given input vector and a pre-defined center vector. There are different types of radial basis function. A normalized Gaussian function usually used as the radial basis function, that is:

$$\emptyset_i(x) = exp \left(\frac{\|x - \mu_i\|^2}{2\sigma_i^2} \right) \qquad (2)$$

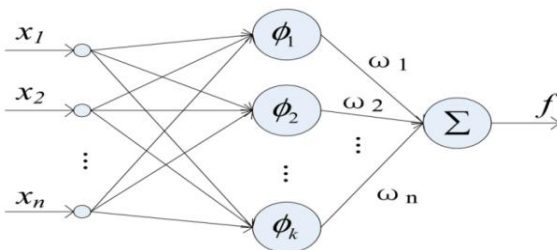Where μ_i and σ_i denote the center and spread width of the i-*th* node, respectively.

Fig.5: The structure of a RBFNN

3.4.2 RBFNN Parameters

In RBF network experiments, two parameters were varied: radial basis function spread and minimum squared error (MSE). The number of neurons in the hidden layer was chosen based on the MSE by the following rule: neurons are added in the hidden layers until it meets the specified MSE. The hold-out method was used with training and testing phase, where the training and testing data set size of each was 50% of total data size [16]. The best configuration of the RBF in this study is shown in table.1.

Table.1 RBF network parameters

| Feature Extraction Technique	Input nodes	Hidden Nodes	Spread Parameter value
GLCM features	5	20	10
FDCT features	72	20	17

4. RESULTS

Performance of the RBF neural network is represented in Fig.6 for both GLCM based features and in FDCT based features. In GLCM based features Experiment, 23 of the 25 benign lung nodules were correctly identified as benign (TN), 2 nodules were misclassified as malignant (FP) as shown Table 2. On the other hand, 22 of the 25 malignant lung nodules were identified as malignant (TP) and 3 nodules were misclassified as benign (FN). In FDCT based features Experiment, 24 of the 25 benign lung nodules were correctly identified as benign (TN), only 1 nodule was misclassified as malignant (FP). On the other hand, 24 of the 25 malignant lung nodules were identified as malignant (TP) and only 1 nodule was misclassified as benign (FN) as shown Table 3. The results of classification obtained using the two different sets of extracted features are depicted in table 4.

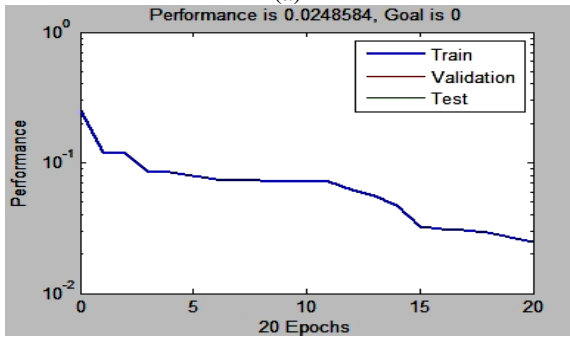**Table.2 Confusion Matrix of Obtained Results from Test Set using GLCM features**

CAD System	Positive	Negative
Positive(Malignant)	22	3
Negative(Benign)	2	23

Table.3 Confusion Matrix of Obtained Results from Test Set using FDCT features

CAD System	Positive	Negative
Positive(Malignant)	24	1
Negative(Benign)	1	24
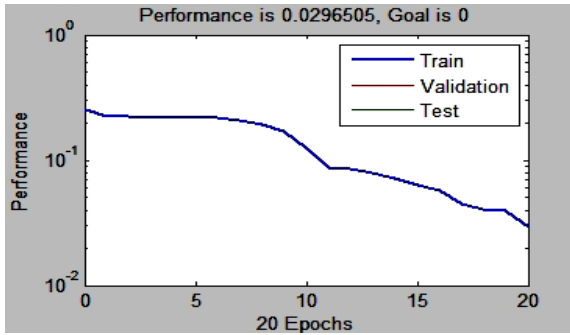
(a)

(b)

Fig.6: (a): Performance of RBF using GLCM based features, (b): Performance of RBF using FDCT based features

(a)

(b)

Fig.7: (a): ROC curve of RBF using GLCM based features, (b): ROC curve of RBF using FDCT based features

To evaluate the performance of the proposed system, the following performance measures were used:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FB} \times 100 \qquad (3)$$

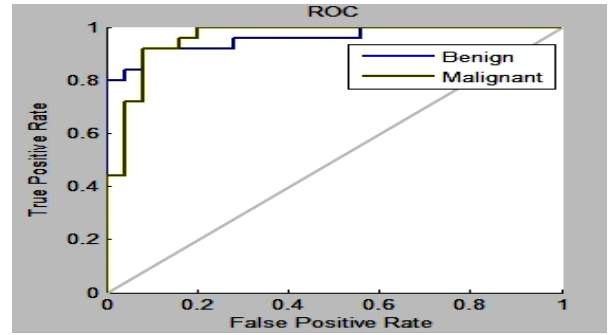$$Sensitivity \frac{TP}{TP + FN} \times 100 \qquad (4)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \qquad (5)$$

Where *TP, TN, FN, FP* stands for true positive, true Negative, false positive and false negative respectively.
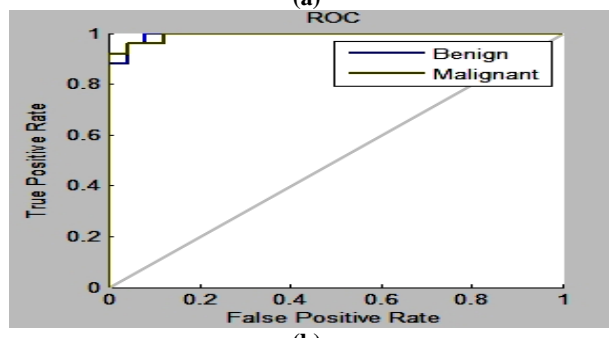
Table.4 Performance Measures

Feature Extraction Technique	Accuracy	Sensitivity	Specificity
GLCM features	90%	88%	92%
FDCT features	96%	96%	**96%**

The plot of sensitivity versus specificity is called receiver operating characteristic (ROC) curve. As concerning statistical indicators of the ROC curve, the primary statistic derived from the ROC is the area under the curve (AUC). The total area under the ROC curve is a measure of the overall performance of a diagnostic test; the larger the area, the better the performance [17]. Fig.7 presents the ROC curve of the system obtained for classification accuracy in the designed CAD system. As seen in Fig.8 in the right, the area under the ROC curve is larger than that in Fig.8 in the left, which gives the FDCT a better classification rate than that of the GLCM.

5. CONCLUSION

In this study, an automatic CAD system that successfully differentiates the lung tumor as benign or malignant on CT images is presented. The system consists of four main parts: image preprocessing, segmentation of Lung region and ROI, feature extraction and selection, and neural network classifier. The segmentation process is done using a set of morphological operations and adaptive threshold method. Two sets of features were extracted using two different features extraction techniques. These are: GLCM based features and FDCT based features. The Classification is done using radial basis function neural network. A correct classification rate of 96% is achieved using FDCT based features which is higher than the correct classification rate using GLCM based features by 4%.

6. REFRENCES

[1] American Cancer Society, Cancer Facts and Figureures 2014. Atlanta, GA:American Cancer Society, 2014

[2] H. Fujita S. Itoh Y. Lee, T. Hara and T. Ishigaki. "Automated detection of pulmonary nodules in helical ct images based on an improved template-matching technique". *IEEE* Transactions on Medical Imaging, 20, 2001.

[3] M. Prokop I. Sluimer, A. Schilham and B. van Ginneken. "Computer analysis of computed tomography scans of the lung: A survey". *IEEE* Transactions on Medical Imaging, 25(4):385–405, April 2006.

[4] E.Dandil, M. Cakiroglu, Z.Eksi, M.ozkan, Oz. Kar Kurt, Ar. Canan, "Artificial Neural Network-Based Classification System for Lung Nodules on Computed Tomography Scans", International Conference of Soft Computing and Pattern Recognition, 978-1-4799-5934-1/14/$31.00 ©2014 *IEEE*

[5] Yongjun WU, Na Wang, Hongsheng Zhang, Lijuan Qin, Zhen YAN, Yiming WU," Application of Artificial Neural Networks in the Diagnosis of Lung Cancer by Computed Tomography", Sixth International Conference on Natural Computation,2010.

[6] J. Kuruvilla, K. Gunavathi, "Lung cancer classification using neural networksfor CT images", Computer Methods and Programs in Biomedicine, vol. 113, pp. 202–209, 2014.

[7] Amal Farag, Asem Ali, James Graham, Aly Farag, Salwa Elshazly and Robert Falk, "Evaluation of geometric feature descriptors for detection and classification of Lung Nodules in low dose CT scans of the chest" ,*IEEE* transactions, PP.169-172, 2011.

[8] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs " Medical Image Analysis vol. 12, pp. 240-258, 2008.

[9] Lung Image Database Consortium (LIDC): https://imaging.nci.nih.gov/ncia/login.jsf

[10] Tinku A., Ajoy K., "Image Processing Principles and applications," John Wiley & Sons, Inc., Hoboken, New Jersey, 2005

[11] Sasidhar B, Ramesh Babu D R, Bhaskarao N, Basheer Jan, "Automated Segmentation of Lung Regions and Detection of Lung Cancer in CT scan", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-4, April 2013.

[12] Randy Crane, A simplied approach to Image Processing: classical and modern technique in C. Prentice Hall, ISBN0-13-226616-1, 1997.

[13] A. Eleyan and H. Demirel, "Co-occurrence matrix and its statistical features as a new approach for face recognition", Turk J Elec Eng & Comp Sci., Vol.19, No.1, pp. 97-107, 2011.

[14] E.J. Candes, L. Demanet, D.L. Donoho, L. Ying, "Fast Discrete Curvelet Transforms" Technical Report, Cal Tech, 2005.

[15] Brijesh K. Verma, Handwritten Hindi Character Recognition Using Multilayer Perceptron and RBF Neural Networks, *IEEE Trans. On Neural Networks*, Vol.21, August 2004, pp. 2111-2115.

[16] Hanan M. Amer, Fatma E.Z. Abou-Chadi, and Marwa I. Obayya,"A Computer-Aided System for Classifying Computed Tomographic (CT) Lung Images Using Artificial Neural Network and Data Fusion", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.10, October 2011.

[17] Sebastian PINTEA & Ramona Molodovan "The receiver-operating characteristic (roc) analysis: fundamentals and applications in clinical psychology" Journal of Cognitive and Behavioral Psychotherapies, Vol. 9, No. 1, March 2009, 49-66.