SCREENING LUNG CANCER IN CT IMAGES USING ARTIFICIAL NEURAL NETWORK

N.Rabecca ,S.Gayathri ,M.Preethi monika
UG Scholars of ECE Department
DMI College of Engineering

A.Milton Jesu Rajan
Assistant Professor of ECE
DMI College of Engineering

Abstract— Lung cancer is one of the most dangerous cancer type in the world. Early detection can save the life and increase the survivability of patients. In this project we obtain a solution for lung cancer symptom detection by applying Shape based image retrieval (SBIR). Our algorithm is broadly divided into three parts, at first part we accept the data set of cancer symptoms which is a generalized way for creating the patterns for Lung Cancer Framework, and in the second part we find the relevant data from the patterns using segmentation approach. We can choose the frequent symptoms only by using the threshold value. Based on the threshold value we decide whether it's a cancer cell or non-cancer cell. We initialize the cancer cell value to support the pattern of cancer symptoms. It is updated in each trial. By updating the cancer cell value in each step we can check the symptom precision which either increases the accuracy or decreases it. Finally result analysis can proved by the appropriately using artificial neural network algorithm.

Keywords: Dicom, SVM, ANN, k-NN, Accuracy, Sensitivity, Precision and Specificity.

I.INTRDUCTION

In medical field the target of Content based image retrieval is to allow radiotherapist to retrieve pictures of comparable options that cause similar diagnosis as the input image. This can be completely different from various field wherever the target is to seek out the closest image from the similar category of an image. So CBMIR: Content Based Image Retrieval Medical Images such methods cannot directly be concerned within the medical field.

During this paper, the image retrieval provides a versatile means that of looking an image based on the description of the required image. The foremost acquainted cancer that happens sometimes for men and women is lung cancer. In keeping with the report carcinoma analyzed and 28% for all cancer deaths. The survival rate for lung cancer analyzed in five years is simply 15%. If the disease is known whereas it's still localized, this rate will increases to 49%. Still, only 15% of diagnosed lung cancers are at this primary stage. This cause the necessity of lung nodule detection [1] in chest Computer Tomography (CT) images in advance. There are various detection techniques, that can be either noninvasive biopsy techniques presently used for early lung cancer detection. Some of them are highlighted. The common detection methods for carcinoma are classical imaging methods like chest radiography (film or digital) and computed tomography (CT).

Digital radiography provides higher contrast resolution with equal or higher spatial resolution when put next to classical radiography techniques [2]. However, these techniques still do not give definitive information which will be utilized toward the first detection of tumors. Low-dose spiral/helical CT will be a promising modality for carcinoma screening. However, it is restricted to tiny peripheral lesions. Chain smokers develop tumors situated within the central airways, and as a result, other. Thus, Researchers have become more and more involved with the elaboration of automated CAD systems for lung cancer. Several publications projected completely different automated nodule recognition systems by image processing, and as well as, completely different methods for segmentation, feature extraction and classification. Computer Tomography (CT) has been observed as the most sensitive imaging technique for early detection of carcinoma. There is a demand for automated methodology to form use of enormous quantity of information obtained CT images. Computer Aided Diagnosis (CAD) will be used coherently for early detection of carcinoma. The consumption of existing CAD system for early detection of lung cancer with the assistance of CT images has been dissatisfactory due to its low sensitivity and False Positive Rates (FPR). This study presents a CAD system which may mechanically discover the lung cancer nodules with reduction in false positive rates. During this study, completely different image processing techniques are given initially in order to attain the lung region from the CT scan chest images. Then the segmentation is transferred with the assistance of clustering algorithm. Finally for automatic recognition of cancer nodules, Support Vector Machine (SVM) is employed that helps in higher classification of cancer nodules. The experiment is conducted for the projected technique by CT images. In many articles, content based access to medical pictures for helping clinical decision-making has been projected that might ease the management of clinical data.

II RELATED WORK

The reported CBIR system particularly on lung CT images are ASSERT, BRISC etc. One of the earliest CBIR systems that focuses on lung CT images is the ASSERT project at Purdue University, which was first published in 1999. It investigated image features such as co- 3 occurrence statistics, shape descriptors, Fourier transforms and global gray level statistics. It used nearest-neighbor and multidimensional hashing for calculating similarity and the best precision reported by this system was 76.3%. Kawata et al. [6] developed a CBIR system on lung nodule in 2004 considering shape descriptors and density histograms to classify and retrieve 3-D lung CT volumes but precision and recall of this CBIR system was not reported. Disney et al. [7] in 2007 developed an open source pulmonary nodule image retrieval framework from chest CT images named as BRISC. They used gray-level co-occurrence features, Haralick feature, Gabor filter and achieved precision of 88% when one nodule is retrieved. Automated nodule segmentation interface and shape based feature of nodule were not explored in their studies. Different modules of designing nodule CBIR system are preprocessing, nodule detection, false positive reduction, nodule segmentation, feature extraction, similarity measure and retrieval of similar nodules. The study was performed by acquiring the heat patterns emitted by the object. Similarly, Sarkar et al. [8] developed a CMOS image sensor for the detection of incoming light rays using polarization information. It was shown that the polarimetric information like DOP, electric field vector intensities and some Stokes parameters like ellipticity and azimuthal angle can be used as a reference source for navigation with little complexity retrieval. In [9], the authors affirm that an Adaptive Median filtering is required to correct the poor contrast that occurs due to poor lighting conditions during image acquisition. They generated a low frequency image by changing each of the pixel value with a median pixel value computed over a square area of 5x5 pixels. Then, a contrast limited adaptive histogram (CLAHE) equalization method is used to improve the contrast of the CT pre-processed image. At the same time, Farag et al. insist in [10] that the filtering approach to use must preserve object boundaries and brief structures, Sharpen the discontinuities to enhance morphological structures and efficiently remove noise in homogeneous physical areas. In their work, the authors used both the Wiener and anisotropic diffusion filters.

Recently, various filters have been developed to enhance lung structures in 3-D images. Many researchers employed filters depending on eigen values of the Hessian matrix [11]. Frangi et al. [12] further developed this approach by introducing a 3D multi-scale structure enhancement filter depending on the eigen values of the Hessian matrix and applying it to the enhancement of vessels. Later, Rikxoort et al. was the first to propose a supervised enhancement applied on single phase and multi-phase techniques [13]. In [14], the authors applied a set of 3D morphologic filters to differentiate the nodule from other surroundings structures, like vessels and bronchi. Segmentation of the lung regions is the next stage of the methods processing scheme. It refers to the process of partitioning the pre-processed CT image into various layers to separate the pixels or voxels corresponding to lung tissue from the surrounding anatomy scheme on two CT data sets. An efficient lung nodule extraction scheme with accuracy of 80.36% in [14] is developed by executing nodule segmentation through weighted fuzzy probabilistic method in [15] clustering is carried out for lung cancer images.

III. PROPOSED METHOD

The proposed ANN framework is shown in the following figure (Figure.1). The database, where the images are kept is called Image database. In the preprocessing technique, the images are enhanced, segmented, and subdivided in order to make flexible work surroundings for further processing works. The proposed model is a combination of feature extraction methods of texture and gray scale resolution. Then this combined form of feature set is stored as a single feature vector in the feature database. When the user provides a query image, the same process steps (such as pre-processing, feature extraction steps) are processed as in the offline image database process for obtaining the feature vector value of query image. Then this query image feature vector value will be related with feature vector value of the feature database. Based on the result, images that are closely similar to the query image are retrieved from the databases and displayed

The initial stage of the proposed Computer Aided Diagnosing (CAD) (Wiemker et al., 2003; Wiemkeret al., 2002) techniques is the extraction of lung region from the CT scan image. The basic image processing techniques are utilized for this cause.

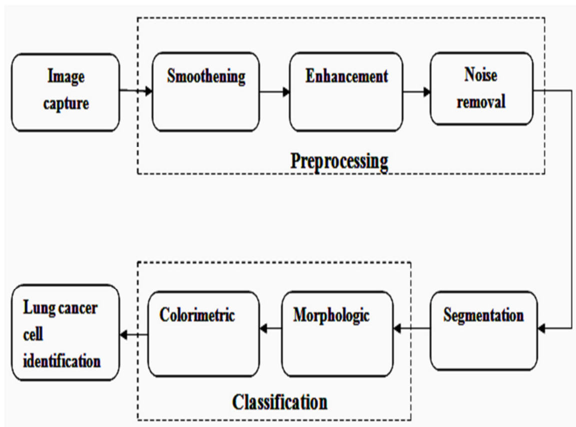

Figure1: Proposed system as a diagnosis

A. Pre processing Enhancement

The lung image will have a significant clarity with post processing enhancement for detecting the nodules. The various steps that have evolved in enhancement after segmentation are given below:

1) Small objects are excluded by morphological opening that is seen inside and outside the lungs in segmented image.
2) Next the borders enhancement and the empty spaces in the border is closed by morphological closing.
3) Morphological operation is followed by canny edge detection that helps to extract the boundary of the enhanced image.
4) Morphological thinning is then carried out with the boundary extracted image.
5) To obtain the last post-processed image morphological filling is done to remove unwanted muscle part from an image except the lungs. Figure 3 shows the post-processing enhancement technique elaborately.

B. Lung CT Image Segmentation

Segmentation of an image consists of separation of the lung nodule from various parts of the CT scan image and then enhancement of the obtained image to get details. This process contains of series of steps that are shown.

1) The input image is changed into grayscale image and Non Local Mean filter helps to remove Gaussian white noise.

2) Otsu's threshold is applied to do segmentation of lung part from the CT image of lungs.

Figure 2 shows the original image, segmented image along with the background eliminated image

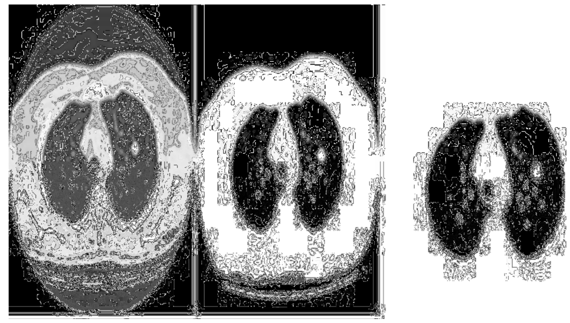

Fig 2: Segmentation

C. Algorithm of ANN

Genetic algorithm provides the optimal solution based on fitness function. Our proposed algorithm includes feeding the GA with an initial selection of population. These populations come from raw dataset of Microarray which represents the components of chromosome. The next step is calculating the fitness function. Based on output of these calculations; the higher fitness could be reserved and discarding the lower one. Third step is the crossover process, that responsible to yield good generation by combine the best component from different genetic. Finally, mutation process is implemented which will generate new gene structure with small random probability. Figure 1 illustrates the proposed system components.

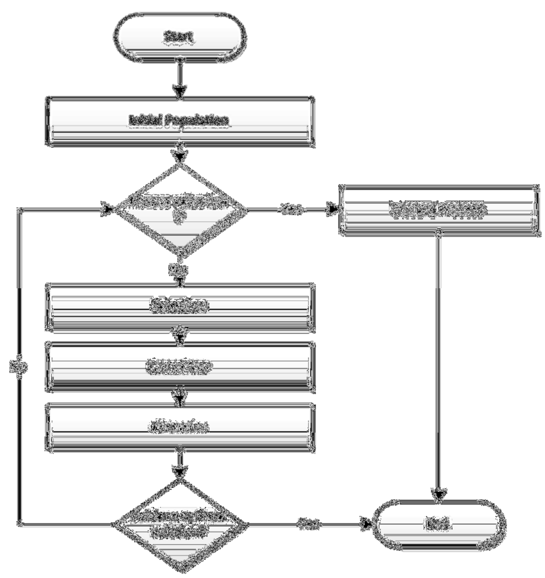

Figure 1. Proposed Genetic Algorithm Components

The previous flowchart shows the general form of the algorithm that will be followed in this study and each stage of this flowchart have own technique. The first is to follow the flowchart so as to get satisfactory results regardless of the used technique. The below steps clarifies the different stages of the proposed GA.

D. Lung Nodule Feature Extraction and Classification

The aim of Feature Extraction is to highlight the required characteristics of the nodule, and it is usually considered to be one of the critical problems of nodule prediction. Highlighting certain features that are vital for the nodule, but excludes the unimportant attributes is the way to structure the nodule. Hence the non-similar features for the lung nodule differentiation are considered and the feature vector thus formulated is FV = {F1, F2, F3, F4, F5, F6}.

3.3.1 Structural Feature

Artificial Neural Network (ANN) is one of the artificial neural networks, which has been proposed for segmenting both gray-level and color images. In[12], the authors present the segmentation problem for gray-level images as minimizing a suitable energy function with ANN, it derived the network architecture from the energy function, and classify the sputum cells into nuclei, cytoplasm and background classes, where the input was the RGB component of the used images. In our work we used the ANN algorithm as our segmentation method. The ANN is very sensitive to intensity variation and it can detect the overlapping cytoplasm classes. ANN is considered as unsupervised learning. Therefore, the network classifies the feature space without teacher based on the compactness of each cluster calculated using the Euclidean distance measure between the k^{th}pixel and the centroid of class l. The neural network structure consists of a grid of N x M neurons with each column representing a cluster and each row representing a pixel. The network is designed to classify the image of N pixels of P features among M classes, such that the assignment of the pixels minimizes the criterion function

$$E = \frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{M} R_{kl}^2 V_{kl}^2 \qquad (1)$$

Where R_{kl} is considered as the Euclidean distance measure between the k th pixel and the centroid of class l, V_{kl} is the output of the k^{th}neurons. The minimization is achieved using ANN and by solving the Motion equation satisfying

$$\frac{\partial u_i}{\partial t} = -\mu(t)\frac{\partial E}{\partial V_i} \qquad (2)$$

Where $\mu(t)$ is as defined in[12] a scalar positive function of time used to increase the convergence speed of the ANN. By applying the relation (2) to equation (1), we get a set of neural dynamics given by:

$$\frac{dU_{kl}}{dt} = -\mu(t)\left[R_{kl}^2 V_{kl}\right] \qquad (3)$$

where U_{kl} and $V_{kl}$are the input and output of the k^{th}neuron respectively. To assign a label m to the k^{th}pixel we use the input-output function given by:

$$V_{km}(t+1) = 1, \text{if} U_{km} = \text{Max}[U_{kl}(t), \forall 1] \quad (4)$$
$$V_{kl}(t) = 0, \text{otherwise.}$$

The ANN segmentation algorithm can be summarized in the following steps:

1. Initialize the input of neurons to random values.
2. Apply the input-output relation given in (4) to obtain the new output value for each neuron, establishing the assignment of pixel to classes.
3. Compute the centroid for each class as follow:

$$\overline{X}_L = \frac{\left[\sum_{k=1}^{n} X_K V_{kl}\right]}{n_l} \qquad (5)$$

Where $n_l$is the number of pixels in class l.

4. Solve the set of differential equation in (3) to update the input of each neuron:

$$U_{kl}(t+1) = U_{kl}(t) + \frac{dU_{kl}}{dt} \qquad (6)$$

5. Repeat from step 2 until convergence then terminate. We applied the ANN with the specification mentioned above to 1000 sputum color images and maintained the results for further processing in the following steps. Our algorithm could segment 97% of the images successfully in nuclei, cytoplasm regions and clear background. Furthermore, ANN took short time to achieve the desired results. By ex-periment, ANN needed less than 120 iterations to reach the desired segmentation result in 36 seconds.

Fuzzy Clustering:

Clustering is the process of dividing the data into homogenous regions based on the similarity of objects; information that is logically similar physically is stored together, in order to increase the efficiency in the database system and to minimize the number of disk access[9]. The process of clustering is to assign the q feature vectors into K clusters, for each k^{th} cluster C^k is its center. Fuzzy Clustering has been used in many fields like pattern recognition and Fuzzy identification. A variety of Fuzzy clustering methods have been proposed and most of them are based upon distance criteria[10]. The most widely used algorithm is the Fuzzy C-Mean algorithm (FCM), it uses reciprocal distance to compute fuzzy weights. This algorithm has as input a predefined number of clusters, which is the k from its name. Means stands for an average location of all the members of particular cluster and the output is a partitioning of k cluster on a set of objects. The objective of the FCM cluster is to minimize the total weighted mean square error[11]:

$$\mathbf{J = Wq^k, C^{(k)}} = \sum_{(\mathbf{q=1,Q})} \sum_{(\mathbf{k=1,K})} (\mathbf{W_{qk}})^{\mathbf{P}} \, \| \mathbf{x^{(q)} - C^{(k)}} \|^{\mathbf{2}} \qquad (7)$$

The FCM allows each feature vector to belong to multiple clusters with various fuzzy membership values. Then the final classification will be according to the maximum weight of the feature vector over all clusters. The detailed algorithm[15]:

Input: Vectors of objects, each object represent s dimensions, where $v = \{v^1, v^2, \ldots\ldots, v^n\}$ in our case will be an image pixels, each pixel has three dimensions RGB, K = number of clusters.

Output = a set of K clusters which minimize the sum of distance error.

1. Initialize random weight for each pixel, it uses fuzz weighting with positive weights $\{W^{qk}\}$between[0, 1].

2. Standardize the initial weights for each qth feature vector over all K clusters via:

$$W^{qk} / \textstyle\sum_{r=1,k} W_{qr} \qquad (8)$$

3. Standardize the weights over $k = 1, \ldots, K$ for each q to obtain W_{qk}, via:

$$W_{qk} \Big/ \sum_{r=1,k} W_{qr} \qquad (9)$$

3. Standardize the weights over $k = 1, \ldots, K$ for each q to obtain W_{qk}, via:

$$W_{qk} = \textstyle\sum_{(r=1,Q)} W_{rk}, q = 1, \ldots\ldots, Q \quad (10)$$

4. Compute new centroids $C(k), k = 1, \ldots., K$ via

$$C^{(K)} = \textstyle\sum_{(q=1,Q)} W_{qk} X^{(q)}, K = 1, \ldots\ldots, K \quad (11)$$

5. Update the weights $\{ W_{qk}, \}$ via

$$W_{qk} = (1/\| x^q - c^k \|^2)^{1/(p-1)}, \quad (12)$$
$$k = 1, \ldots.. K, q = 1, \ldots., Q$$

6. If there is change in the input, repeat from step 3, else terminate.

7. Assign each pixel to a cluster based on the maximum weight. We applied the FCM clustering algorithm with the specification mentioned above to 1000 sputum color images and maintain the results for further processing in the following steps. Our algorithm could segment the images into nuclei, cytoplasm regions and clear background, however, the FCM is not sensitive to intensity variation, therefore, the cytoplasm regions are detected as one cluster when we fixed the cluster number to three, four, five and six. Moreover, FCM failed in detecting the nuclei, it detected only part of it. By experiment, the FCM algorithm takes less than 50 iterations to reach the desired results in 10 seconds on average.

IV. EXPERIMENTAL RESULTS

The input set of the lung pictures considered are taken for National Lung Screening Trial (NLST) data/images of stage I and II. The amounting sample images taken for experimentation are 111 for stage-I and 73 samples are taken for stage-II lung cancer. Out of this four-fifth of the info helps for training and the rest of one-fifth is taken for checking the classifiers.24 pictures of stage I and 17 picture of stage II are in test dataset. Confusion matrix is shown in first table TP is 24, tells 24 images of stage I are predicted as stage I, FP is 2, means 2 images of stage II is expected to be in stage I, FN is 0, shows that no picture of stage I is expected to be in stage II. TN is 15, shows that 15 pictures of stage II are expected to be in stage II.

Then the planned system can process through its processing steps and it will detect whether the supplied lung image is with cancer or not at last. The results prove that there are few miss-detections however overall potency of vision based potency measure is more than 95%.

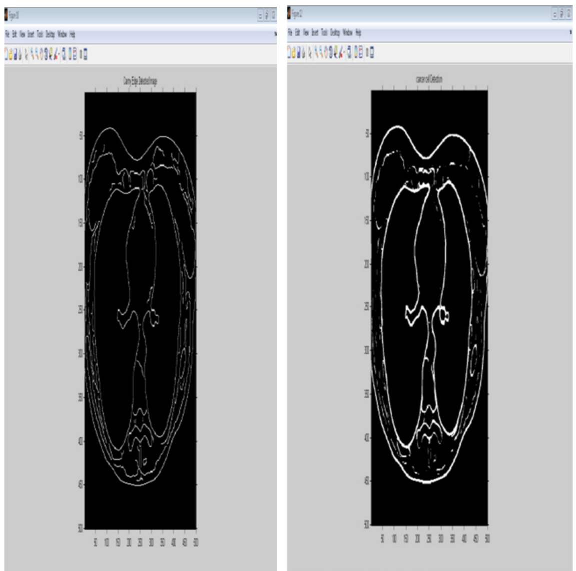

Figure 4: (a) canny edge detection (b) Lung Nodule detected after Segmentation.

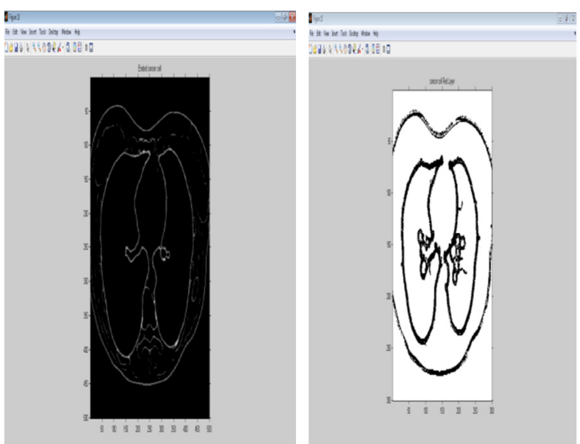

Figure 5: (a) classification (b) disease detection

V. CONCLUSION

The field of disease designation may be ceaselessly evolving and highly active field of analysis. The intention of this study was to predict the standing of patient for early detection of carcinoma. A completely unique approach for predicting carcinoma nodule at early stage by SVM Classifier has been planned here. The Structural and Textural Features have been used for describing the nodule. The result is highly encouraging and the information was tested on SVM Classifier with RBF kernel obtained associate accuracy of 95.12%. A comparison of classification accuracy for ANN, KNN and SVM Classifiers was created on lung CT scan pictures of stage I and stage II. The classification rates obtained for the SVM, ANN and k-NN Classifier were 95.12%, 92.68% and 85.37% for the check pictures. The results show that there are few miss-detections however overall potency of vision based potency measure is over 95%. By achieving this accuracy, lives of many patients with abnormal case of lung cancer can be saved.

REFERENCE

[1] Cancer Facts and Figure 2009 by American Cancer Society, http://www.cancer.org

[2] Anthony V D'Antoni, Genevieve Pinto Zipp, Valerie G Olson and Terrence F Cahill,"Does the mind map learning strategy facilitate information retrieval and critical thinking in medical students?", BMC Med Educ. 2010.

[3] Stefan Diederich et al., "Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers", Radiology, vol. 222, no.3, pp. 773-781, 2002.

[4] Ichiro Yoshino, Masafumi Yamaguchi, Testuzo Tagawa, Seiichi Fukuyama, Toshifumi Kameyama, Atsushi Osoegawa and Yoshihiko Maehara, "Operative results of clinical stage I non-small cell lung cancer", Lung Cancer, vol. 42, no. 11, May 2003.

[5] Austin et al., "Glossary of terms for CT of lungs; recommendations of the Nomenclature Committee of the Fleischner Society", Thoracic Radiology, vol. 200, pp. 327-331, April 1996.

[6] Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Yamada, K. Mori, H. Nishiyama, K. Eguchi, M. Kaneko, and N. Moriyama, "Pulmonary nodule classification based on nodule retrieval from 3-D thoracic CT image database", Medical Image Computing and Computer-Assisted Intervention (MICCAI 2004).

[7] Michael O. Lam, Tim Disney, Daniela S. Raicu, Jacob Furst and David S. Channin, "BRISC-An Open Source Pulmonary Nodule Image Retrieval Framework", Journal of digital imaging, 2007.

[8] Arimura, S. Katsuragawa and K. Suzuki, "Computerized scheme for automated detection of lung nodules in low-dose

computed tomography images for lung cancer screening", Acad. Radiol., Vol. 11, pp. 617629, 2004.

[9] Ambrosini,S. Nicolini, P. Carolia, C. Nannia, A. Massarob, M.-C. Marzolab, D. Rubellob and S. Fantia, "PET/CT imaging in di_erent types of lung cancer: An overview", European Journal of Radiology, Vol. 81, pp. 988-1001, 2013.

[10] El-Bazl, A. Farag, R. Falk and R. LaRocca, Automatic identification of lung abnormalities in chest spiral CT scans, In proc. of the international conference on Acoustics, Speech, and Signal Processing (ICASSP '03), Vol.2, pp. 261-264, 2003.

[11] El-Baz, A. Farag, G. Gimelfarb, R. Falk, M.-A. El-Ghar and T. Eldiasty, "A framework for automatic segmentation of lung nodules from low dose chest CT scans", in Proc. of the 18th International Conference on Pattern Recognition (ICPR 06), Vol. 3, pp. 611614, 2006.

[12] El-Baz, G. Gimelfarb, R. Falk and M. Abo El-Ghar, "3D MGRF-based appearance modelling for robust segmentation of pulmonary nodules in 3D LDCT chest images, in Lung Imaging and Computer Aided Diagnosis", chapter 3, pp. 5163, Taylor and Francis edition, 2011.

[13] Ezoe, H. Takizawa and S. Yamamoto, "An automatic detection method of lung cancers including ground glass opacities from chest X-ray CT images", In Proc. of SPIE, vol. 4684, pp. 16721680, 2002.

[14] Ashwin, S.-A. Kumar, J. Ramesh and K. Gunavathi, "E_cient and Reliable Lung Nodule Detection using a Neural Network Based Computer Aided Diagnosis System", In Proc. of the International Conference on Emerging Trends in Electrical Engineering and Energy Management (ICETEEEM'2012), pp. 135-142, Chennai, 13-15 Dec. 2012.

[15] Farag, J Graham, A. Farag and R. Falk, "Lung Nodule Modelling A Data-Driven approach", Advances in Visual Computing, Vo. 5875, pp 347-356, 2009.

[16] Haussecker and B. Jahne, "A tensor approach for local structure analysis in multidimensional images in 3-D", Image Anal. Synthesis, pp.171178, 1996.

[17] Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering", Med. Image Computing Computer Assisted Intervention, vol. 1496, pp. 130137, 1998

[18] E.-M. vanRikxoort, B. vanGinneken, M. Klik, and M. Prokop, "Supervised Enhancement Filters: Application to Fissure Detection in Chest CT Scans", IEEE trans. on Medical imaging, Vol. 27, No. 1, pp. 1-10, 2008.

[19] D.-F. Yankelevitz, A.-P. Reeves, W.-J. Kostis, B. Zhao and C.-I. Henschke, "Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation", Radiology, Vol. 217, No. 1, pp. 251256, 2000.

[20] K.-Z. Faizal and V. Kavitha, "An E_ective Segmentation Approach for Lung CT Images Using Histogram Thresholding with EMD Refinement", Proceedings of International Conference on Internet Computing and Information Communications, Advances in Intelligent Systems and Computing, Vol. 216, pp. 483-489, 2014.

[21] H. Shao, L. Cao and Y. Liu, "A Detection Approach for Solitary Pulmonary Nodules Based on CT Images", in Proc. of the 2nd International Conference on Computer Science and Network Technology, pp. 1253-1257, 2012.

Zhou, T. Hayashi, T. Hara, H. Fujita, R. Yokoyama, T. Kiryu and H. Hoshi, "Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images", Computerized