

# Lung cancer detection from chest CT images using spatial FCM with level set and neural network classifier.

Manikandan T, Suganya V

Department of ECE, Rajalakshmi Engineering College, Chennai, India

## Abstract

Accurately and reliably automated segmentation of nodule could play an important role in lung cancer diagnosis. The chest Computer Tomography (CT) lung images are used to detect real malignant (cancerous) nodules. An effective Spatial Fuzzy C-means clustering with level set is proposed in this work to effectively segment the suspected lung nodules from CT images in order to detect the lung cancer. After segmentation, features were extracted and fed to neural network for classification. The classification process is done by using feed forward-back propagation in neural network. Performance of the proposed system was evaluated using 106 subjects' Computed Tomography (CT) images retrospectively obtained from the Bharat Scans, Chennai. The proposed method reduced the false positive nodule candidates significantly. It has achieved the sensitivity and accuracy of 88% and 84%, respectively.

**Keywords:** Spatial fuzzy C-means, Nodules, Neural network, Classification.

*Accepted on February 11, 2017*

## Introduction

Lung cancer is one of the crucial cancers to cure and the mortality rate of lung cancer is huge among all other types of cancer [1]. One of the most serious cancers in the world is lung cancer, with a gradual increase in number of deaths every year [2]. Detection of lung cancer in its early stage can be helpful for medical treatment to limit the danger [3]. Estimated 85% of lung cancer cases in male and 75% in female are caused by smoking. In the earlier stage, lung cancer generally exists in the form of nodule. The size of nodule and its growth rate is an important criterion to classify between the malignant and benign nodules. Tumors can be benign or malignant; the word tumor refers to malignant. Benign tumor can be removed usually on the other hand malignant tumor grow aggressively. The processes of spreading termed metastasis, the area of growth at these distant sites are metastases.

The diagnosis of lung cancer at early stage is critical and uncertain as the physicians direct the patient to undergo biopsy only after between time intervals of 6 to 18 months. The advanced imaging techniques such as Computed Tomography (CT) scanning that precisely capture the images of lung are available, finding the cancerous nodules is still a challenging task for physicians [4,5]. CT scan of lung produces continuous cross-sectional images and to confirm the cancerous nature of lung, it is essential to analyze every cross section.

Lung Cancer is diagnosed generally by analysing a tissue cluster formation called 'Nodule' inside the lung [6]. Based on the shape, the nodules are classified as Well-Circumscribed, Pleural tail, Vascularised and Juxta Pleural. Well Circumscribed nodules are placed well inside the Lung parenchyma region, Pleural Tail nodules are connected to the pleural surface by a thin tail like structure, Vascularised nodules are attached with the vessels present inside the lung region, and the Juxta Pleural nodules are connected significantly to the wall surface inside the lung. The shape and size of the Nodules are varying; the process

of segmentation is really a challenging task. The nodules segmented initially based on its shape, size, texture or position is only suspected nodules. Fuzzy clustering is one of the most commonly used algorithms for image segmentation]. Based on fuzzy set theory, the data clustering method is very effective in addressing fuzzy and uncertain problem in gray images. The conventional fuzzy clustering algorithm was not taken into account spatial information when used for image segmentation, sometime leading to unexpected results. Developed an effective algorithm Fuzzy auto-seed cluster means morphological to segment the lung nodules from the consecutive slices of CT image to detect lung cancer.

A nodule detection system consists of three steps: lung segmentation, nodule detection and false positive reduction [7,8]. Optimal thresholding is used for segmentation of lung nodule [9]. A fixed threshold value has been used to segment the lung region after thresholding; lung volume can be extracted from the segmented images seed point in initial region and labelling techniques to segment the lung volume and extracted lung volume need to be refined [10,11]. Due to complexity of these approaches several methods have been presented for refining a lung mask. Morphological dilation operators have been used to include Juxta-pleural nodules and a rolling ball algorithm applied for effective lung mask correction.

The segmented lung volume, nodule have been detected using various methods [12,13]. Multiple gray level thresholds have been applied to lung region to identify nodules. Template matching based method proposed a novel approach based on genetic algorithm (GA) template matching (GATM) technique for detecting nodules within lung area [14]. The shape-based approach also used to detect nodules with spherical elements. Template matching scheme utilize three dimension of CT region, the 3D template being used to find structures with properties similar to nodules.

Another approach used in lung region extraction process based

on pixel classification [15], each pixel in the CT image is classified into anatomical class. Classifier is various types of neural networks training with a variety of local feature including intensity, location and texture measures. The model-based detection approaches, the relatively compact shape of a small lung nodule is taken into account while establishing the models to identify nodules in the lungs. Nodule candidates are detected using template matching in which edge pixels for circles that could cause these edges. After segmentation different features should be extracted for diagnosis between true and false cancerous candidates.

Lung nodules are the indicator of cancer and are the tiny mass inside the lung. Most of the lung nodules are non-cancerous (benign), but about 40% of them are cancerous. Therefore, it is a real challenge for the researchers to quantify and describe the lung nodules. Hence it is essential to develop an efficient algorithm to make an accurate decision on the cancerous nature of the nodules in the initial stage. This work proposed novel algorithm called spatial fuzzy clustering with level set, which can detect the lung cancer in the early stage.

### Proposed Methodology

The proposed methodology shown in Figure 1 involves the following stages: a) Preprocessing, b) Segmentation, c) Feature extraction, and d) Classification.

#### Database

The foremost step in medical image processing is image acquisition. The proposed CAD system used as an input a set of CT scans to be analyzed in order to classify lung nodules. The CT images of those subjects were retrospectively collected from Bharat Scans, Chennai. Thus, the study population comprises of 106 (24 stage-I, 32 stage II and 50 normal) subjects, for whom the analysis was carried out. Two CT chest radiologists were allotted by the department of radiology of Bharat Scans to analyse all the 106 subjects CT scan images, which showed a total of 56 malignant nodules and 745 benign nodules with size varies between 3 mm and 30 mm (used as ground truth images).

#### Preprocessing

Preprocessing is performed on image to improve the quality to increase the precision and accuracy of processing take place

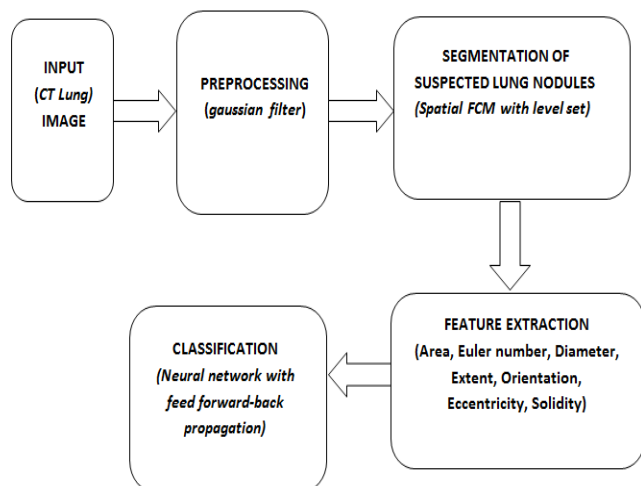


Figure 1. Block diagram of the proposed method.

after this stage [16]. Gaussian filter is used for preprocessing to remove and PSNR value is calculated.

### Segmentation

In Spatial fuzzy clustering with level set, the centroid and the scope of each subclass are estimated adaptively in order to minimize a pre-defined cost function. It is appropriate to take fuzzy clustering as a kind of adaptive thresholding. Fuzzy c-Means (FCM) is one of most popular algorithms in fuzzy clustering, and has been widely used in medical problems [17,18]. The classical FCM algorithm originates from k-means algorithm. In detail, the k-means algorithm seeks to assign N objects, based on their attributes, into K clusters ( $K \leq N$ ). For medical image segmentation, N equals the number of image pixels  $N_x \times N_y$ . The desired results include the centroid of each cluster and the affiliations of N objects Standard k-means clustering attempts to minimize the cost function,

$$J = \sum_{m=1}^N \sum_{n=1}^C \mu_{mn}^1 \|i_n - v_m\|^2 \tag{1}$$

Where  $i_n$  is the specific image pixel,  $v_m$  is the centroid of  $m^{\text{th}}$  cluster, and  $\|\cdot\|$  denotes the norm. The ideal results of a k-means algorithm maximize the inter-cluster variations, but minimize the intra-cluster ones. In k-means clustering, every object is limited to one and only one of K clusters. In contrast, an FCM utilizes a membership function  $\mu_{mn}$  to indicate the degree of membership of the  $n^{\text{th}}$  object to the  $m^{\text{th}}$  cluster, which is defensible for medical image segmentation as physiological tissues are usually not homogeneous. The cost function in an FCM is similar to Eq. (2).

$$J = \sum_{m=1}^N \sum_{n=1}^C \mu_{mn}^l \|i_n - v_m\|^2 \tag{2}$$

Where  $l (>1)$  is a parameter controlling the fuzziness of the resultant segmentation. The membership functions are subject to the following constraints:

$$\sum_{m=1}^C \mu_{mn} = 1; 0 \leq \mu_{mn} \leq 1; \sum_{n=1}^N \mu_{mn} > 0 \tag{3}$$

The membership functions  $\mu_{mn}$  and the centroids  $v_m$  are updated iteratively

$$\mu_{mn} = \frac{\|i_n - v_m\|^{-2/(l-1)}}{\sum_{k=1}^C \|i_n - v_k\|^{-2/(l-1)}} \tag{4}$$

$$v_i = \frac{\sum_{n=1}^N \mu_{mn}^l i_n}{\sum_{n=1}^N \mu_{mn}^l} \tag{5}$$

Assigned high membership values, while those are far away are assigned low values. One of the problems of standard FCM algorithms in segmentation is the lack of spatial information. Since image noise often impairs the performance of FCM segmentation, it would be attractive to incorporate spatial information into an FCM proposed a generalized FCM algorithm that adopts a similarity factor to incorporate local intensity and spatial information. In contrast to the above preparatory weighting, it is also possible to utilize morphological operations to apply spatial restrictions at the post-processing stage incorporated into fuzzy membership functions directly using

$$\mu_{mn}^1 = \frac{\mu_{mn}^p h_{mn}^q}{\sum_{k=1}^c \mu_{kn}^p h_{kn}^q} \quad (6)$$

where p and q are two parameters controlling the respective contribution. The variable  $h_{mn}$  includes spatial information by

$$h_{mn} = \sum_{k \in N_n} \mu_{nk} \quad (7)$$

Where  $N_n$  denotes a local window centred on the image pixel n. The weighted  $\mu_{mn}$  and the centroid  $vm$  are updated as usual according to Equations (6) and (7).

Both FCM algorithms and level set methods are general-purpose computational models that can be applied to problems of any dimension [19,20]. However, if we constrain them to medical image segmentation, it is possible to take advantage of the specific circumstances for better performance. A new fuzzy level set algorithm is thereby proposed for automated medical image segmentation. It begins with spatial fuzzy clustering, whose results are utilized to initiate level set segmentation, estimate controlling parameters and regularize level set evolution. The new fuzzy level set algorithm automates the initialization and parameter configuration of the level set segmentation, using spatial fuzzy clustering. It employs an FCM with spatial restrictions to determine the approximate contours of interest in a medical image. Benefitting from the flexible initialization as in Eq. (4.19), the enhanced level set function can accommodate FCM results directly for evolution. The component of interest in an FCM results is convenient to initiate the level set function as:

$$f_0(x, y) = -4 \in (0.5 - B_k) \quad (8)$$

Where  $\epsilon$  is a constant regulating the Dirac function. The Dirac function is then defined as follows:

$$0, |x| > \epsilon$$

$$\delta_\epsilon(x) = \left\{ \frac{1}{2\epsilon} \left[ 1 + \cos\left(\frac{\pi x}{\epsilon}\right) \right], |x| \leq \epsilon \right. \quad (9)$$

$B_k$  is a binary image obtained from  $B_k = R_k \geq b_0$

Where  $b_0$  ( $\in(0, 1)$ ) is an adjustable threshold. Benefitted from spatial fuzzy clustering,  $B_k$  can in some sense approximate the component of interest, which can be readily adjusted by  $b_0$ . There are several controlling parameters associated with level set methods are important for medical image segmentation. It is therefore necessary to configure them appropriately, which unfortunately differ from case to case. Currently there are merely a few general rules to guide the configuration of these parameters. For example, it is known that a larger  $\sigma$  leads to a smoother image, but sacrifices an image detail. A larger time step  $\tau$  may accelerate level set evolution, but incurs the risk of boundary leakage. Moreover it is necessary to choose a positive  $v$  if the initial  $f_0$  is outside the component of interest and vice versa.

### Feature extraction

The suspected nodules are segmented using Spatial FCM with level set algorithm. Approximately 60% of detected nodules in the lung CT scan would not be cancerous nodule. It is very important not to make the patient panic by categorizing all the nodules as an indicator for cancer. Therefore finding the discriminative mathematical description for the cancerous

and noncancerous nodules is a challenging research area. The structural and Texture pattern of the cancerous and noncancerous nodules need to be mathematically analyzed, and quantitatively specified, so that we can classify the nodules based on their cancerous nature. This Pattern Recognition (PR) process has two steps, Feature Extraction and Classification. The two main categories of features are Structural and Texture features. Structural features can be categorized as Shape and Size features.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. The input data should be transformed into a reduced representation of set of features such as Area Bounding box, Centroid, Eccentricity, Euler number and Diameter. To extract these features of the CT image, region properties are applied.

These features are explained below.

**Area:** The actual number of pixels in the region.

$$\sum_{x=1}^m \sum_{y=1}^n f_b(x, y) \quad (10)$$

**Bounding box:** The smallest rectangle containing the region, a 1-by-Q \*2 vector, where Q is the number of image dimensions.

**Centroid:** 1-by-Q vector that specifies the centre of mass of region.

**Eccentricity:** Scalar that specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. (0 and 1 are degenerate cases; an ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment.) This property is supported only for 2-D input label matrices.

$$Eccentricity = \left( \frac{\max \sum_{x=1}^m f_b(x, ;)}{\max \sum_{y=1}^n f_b(, y)} \right) \quad (11)$$

**Euler number:** Scalar that specifies the number of objects in region minus the number of holes in those objects. This property is supported only for 2-D input label matrices.

**Equivalent diameter:** Scalar that specifies the diameter of a circle with the same area as the region. This property is supported only for 2-D input label matrices.

$$ED = \frac{2}{\pi} \sqrt{area} \quad (12)$$

**Orientation:** Angle in degrees between x axis and major axis of ellipse.

**Solidity:** Proportion of pixels in the convex hull that are also in the region.

$$Extent = \frac{Area}{Area \text{ of Boundingbox}} \quad (13)$$

**Extent:** Specifies the ratio of pixels in the region to pixels in total bounding box.

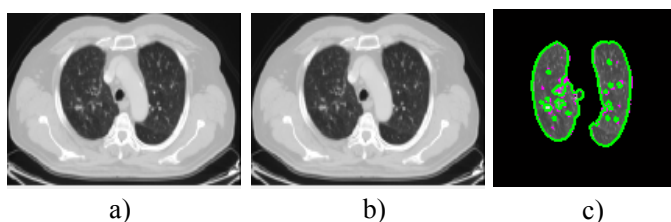
$$Extent = \frac{Area}{Area\ of\ Bounding\ box} \quad (14)$$

**Classification**

In feed forward neural networks information always moves in one direction only, there is no feedback. The information moves forward from input layer through hidden layer to the output layer. The networks used are Hebb, Perceptron, Adaline and Madaline networks. In Hebb network learning is done by modification of the weights of the neurons. The weight is information used by neural network to solve a problem. The Perceptron network is supervised classifier for classifying an input into one of two possible outputs. It is a type of linear classifier. The classification algorithm makes its predictions based on a linear predictor function combining a set of weights with the feature vector describing a given input. Both bias and threshold are needed in this network. The Adaline (Adaptive Linear Neuron) network uses bipolar activations for its input signals and target output. The weights on the connection from the input units to the adaline network are adjustable. The network has a bias, which acts like an adjustable weight on a connection from a unit whose activation is always 1. A Madaline network consists of adalines arranged in a multi-layer net. It is a two layer neural network with a set of adalines in parallel as its input layer and a single processing element in its output layer. The back propagation is a systematic method of training multilayer neural networks in a supervised manner. The backpropagation method, also known as the error back propagation algorithm, is based on the error-correction learning rule. The back propagation network consists of atleast three layers of units: an input layer, at least one intermediate hidden layer and one output layer. The units are connected in feed forward fashion with inputs units connected to the hidden layer units and the hidden layer units are connected to the output layer units. An input pattern is forwarded to the output through input to hidden and hidden to output weights. The output of the network is the classification decision.

**Result and Discussion**

Performance of the proposed CAD system is evaluated using the 106 CT images of Bharat Scans, Chennai. The input CT image is preprocessed using Gaussian filter and then segmented using Spatial FCM with level set. Figure 2a shows the steps involved in the segmentation of suspected lung nodules. After the segmentation process features were extracted. The extracted features used for classification. The preprocessing carried out with Gaussian filter which reduced the noise present in the input image.



**Figure 2.** Steps involved in the segmentation of suspected lung nodules. a) Input image, b) Pre-processed (Gaussian filtered) image and c) Segmented lung and suspected lung nodules.

The first into grayscale image and it nodule detection is carried resized to 255 × 255. Nodule detection is carried out using the following steps. In preprocessing Gaussian filter is used to remove noise and the PSNR (Peak Signal to Noise Ratio) is calculated 34.831dB. The Gaussian filtered image is shown in Figure 2b. The preprocessed image is subjected to Spatial FCM. The filtered image is segmented based on the cluster centres and Segmented image Figure 2c is obtained. After segmentation by applying region properties different features are extracted. The features extracted are Area, Centroid, Euler Number, Eccentricity, Minor Axis, Major Axis, Extent, Solidity and Orientation is extracted. ANN is trained to detect whether the suspected lung nodule is normal or malignant. The proposed system was tested with 106 subjects’ (50 cancerous and 50 normal) Computed Tomography (CT) images, which were retrospectively collected from Bharat Scans, Chennai. Those CT images were analysed by two radiologists of Bharat Scans, which have shown 50 malignant (cancerous) nodules and 745 benign nodules (non-cancerous) with sizes varying between 3 mm to 30 mm. The performance of classifier was analysed using True Positive (TP), True Negative (TN), False Positive (FP) and True negative (TN) are shown in Table 1 and accuracy of classification are shown in Table 2.

**Conclusion**

A computer-aided segmentation and classification method is proposed. The performance of the proposed system was evaluated on 106 CT images. Spatial FCM with level set were used for segmentation and classification was done by neural network. The region of interest is the segmented single slices containing two lungs. The input CT image was pre-processed using Gaussian filter and PSNR value was calculated. Spatial FCM with level set was utilized to segment the preprocessed image. Features were extracted from segmented image and fed to neural network for classification. The feature vectors were classified into malignant nodules and non-malignant (benign) nodules by using the neural network classifier. The proposed work successfully detected lung malignant nodules in 3 mm to 30 mm (initial stage) in size. Thus it will helpful in identifying lung cancer at early stage (Stage I and II). The proposed CAD system has achieved a sensitivity of 88% and accuracy of 84% in the initial stage of lung cancer detection. Therefore physicians can use the proposed CAD system as a decision support system to make the final decision in the initial stage itself. Therefore high possibility of increasing the patient’s survival with lung cancer. The limitation of the work is very low specificity (79%).

**Table 1.** Result of performance measures.

Error rate	Neural Network
TP	44
FN	6
FP	159
TN	586

**Table 2.** Results of classification accuracy.

Neural Network Classifier Accuracy	
Sensitivity	88%
Specificity	79%
Accuracy	84%

## References

1. Manikandan T, Bharathi N. 'Design of lung cancer specific questionnaires for the south Indian population'. *Int J Manag IT Eng*. 2014;4(12):311-31.
2. Manikandan T, Bharathi N. 'A survey on computer-aided diagnosis systems for lung cancer detection'. *Int Res J Eng Technol*. 2016;3(5):1562-70.
3. Manikandan T, Bharathi N. 'A novel semi-automated 3-D CAD visualization system as an aid for surgical planning of lung cancer'. *ARPJ Eng Appl Sci*. 2015;10(4):1872-78.
4. Manikandan T, Bharathi N. 'Lung cancer detection by automatic region growing with morphological masking and neural network classifier'. *Asian Journal of Information Technology*. 2016;15(21):4189-94.
5. Manikandan T, Bharathi N. 'Hybrid neuro-fuzzy system for prediction of stages of lung cancer based on the observed symptom values'. *Biomed Res J*. 2017;28(2):588-93.
6. Manikandan T, Bharathi N. 'Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier'. *J Med Sys*. 2016;40(7):1-9.
7. Samuel G, Saravanan V, Devi MV. Lung Nodule Diagnosis from CT images using Fuzzy Logic. *International Conference on Computational Intelligence and Multimedia Application*.
8. Siegel R, Naishandhan D, Jemal A. Cancer Statistics. *A Cancer Journal for Dinicians*. 62(2012):10-29.
9. Suarez Cuenca J, Tahoces P, Souto M, et al. Application of the Iris Filter for Automatic Detection of Pulmonary Nodules on Computed Tomography Images, *Computers in Biology and Medicine*. 39(2009): 921-23.
10. Messay T, Hardie R, Rogess S. A new computationally efficient CAD system for pulmonary Nodule Detection in CT imagery, *Medical Image Analysis*. 14(2010):309-406.
11. Pectico A, Delogu P, Fantacci M, et al. Lung Nodule Detection in Low-dose and Thin-slice Computed Tomography. *Computers in Biology and Medicine*. 38(2008):525-34.
12. Ozeke S, Camurcer A. Automatic Lung Nodule Detection using Template matching, *Lecture notes in computer Science*. 4243(2006):247.
13. Takizawa H, Yamamoto S, Shiina T. Accuracy improvement of Pulmonary Nodule Detection based on Spatial Statistical information and systems. 90(2007):1168-74.
14. Lee Y, Hala T, Fujita H, et al. Automated detection of Pulmonary Nodules in Helical CT images based on an improved template-matching technique, *IEEE Transaction on Medical Imaging*. 20(2001):595-604.
15. Ginnekan BV, Romeny BM, Viergever MA. Computer-aided diagnosis in chest radiography, *IEEE transaction on medical imaging*. 2012;20(12).
16. Cardles V, Kimmel R, Sapiro G. Geodesic active contours. *International Journal of Computer Vision*. 22(1997):61-79.
17. Chuang KS, Hsiao HL, Chen S, et al. Fuzzy C-means clustering with Spatial information for image segmentation. *Computerized Medical Imaging and Graphics*. 30(2006):9-15.
18. Cai W, Chen S, Zhang D. Fast and Robust Fuzzy C-means Clustering Algorithm incorporation Local information for image segmentation. *Pattern Recognition*. 40(2007).
19. Leiw K, Li BN, Dong MC, Vai MI. AFC-ECG: An Adaptive Fuzzy ECG Classifier, in: *Proceeding of the 11<sup>th</sup> World Congress on Soft Computing in industrial Application (WSCII)*. *Advance in Soft Computing*. 39(2007):189-909.
20. Osher S, Fedkiw R. *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York. 2003.

### \*Correspondence to:

Manikandan T  
 Department of ECE  
 Rajalakshmi Engineering College  
 Chennai, Tamil Nadu  
 India  
 E-mail: manikandan.t@rajalakshmi.edu.in