

Segmentation of CT Lung Images Using FCM with Active Contour and CNN Classifier

Malathi M^{1*}, Sinthia P², Madhanlal U³, Mahendran K⁴, Nalini M⁵

Abstract

Objective: Lung cancer is one of the unsafe diseases for human which reduces the patient life time. Generally, most of the lung cancers are identified after it has been spread into the lung parts and moreover it is difficult to find the lung cancer at the early stage. It requires radiologist and special doctors to find the tumoral tissue of the lung cancer. For this reason, the recommended work helps to segment the tumoral tissue of CT lung image in an effective way. **Methods:** The research work uses hybrid segmentation technique to separate the lung cancer cells to diagnose the lung tumour. It is a technique which combines active contour along with Fuzzy c means to diagnose the tumoral tissue. Further the segmented portion was trained by Convolutional Neural Network (CNN) in order to classify the segmented region as normal or abnormal. **Results:** The evaluation of the proposed method was done by analyzing the results of test image with the ground truth image. Finally, the results of the implemented technique provided good accuracy, Peak signal to noise ratio (PSNR), Mean Square Error (MSE) value. In future the other techniques can be utilized to improve the details before segmentation. The proposed work provides 96.67 % accuracy. **Conclusion:** Hybrid segmentation technique involves several steps like preprocessing, binarization, thresholding, segmentation and feature extraction using GLCM.

Keywords: Lung cancer- Computer Tomography (CT)- Convolutional Neural Network (CNN) - FCM

Asian Pac J Cancer Prev, **23 (3)**, 905-910

Introduction

The lungs are the most important part of respiratory tract. The upper respiratory tract and the lower respiratory tracts are the two tracks of the lungs. Each level of the body requires oxygen to live a healthy life. The lungs are spongy, pinkish organs located at the two upside-down cones in our chest. There are three lobes on right lung and two lobes on left lung to build space for our heart. The lungs start at the bottom of the trachea (wind pipe). It has the ability to carry the air in and out of lungs. The lungs start at the underneath of the trachea (wind pipe). It has the ability to carry the air inside and away from the lungs. Bronchus helps to join the trachea with the lungs. The bronchi airways and trachea appear from an upside-down “y” in your chest. It is named as bronchial tree. The bronchi have many lesser bronchi and the less significant tubes called bronchioles. Like a tree, these miniature tubes extend out into each branch of the lungs. Few of them are so little-like that they have the thickness of a hair. Our body has 30,000 bronchioles in every lung. The bronchiole tube trimmings with a group of undersized air sacs called alveoli (referred as alveolus). There are 600 million of alveoli in our lungs. Mouth and nose nasal

cavity, throat (pharynx) and vector box(larynx) are the parts of upper respiratory tract. The lower respiratory tract has lungs, trachea (wind pipe), bronchi, bronchioles and alveoli. The major purpose of the lungs is to acquire oxygen from the environment and transmit it to the bloodstream. There are many types of lung disorders among which cancer is one of the diseases that affects the lungs. It may be malignant or benign. Cigarette smoking is the main reason for lung cancers. The blood in the human body travels from the heart through lungs. So, there is a chance for easy spreading of cancer to the other parts of the body. Early diagnosis of lung cancer helps to increase the lifetime of the patient.

CT scan imaging is one of the most excellent imaging expertise to identify the tumoral tissue from the captured image. Hence the author proposed computer aided diagnosis in order to identify the Malignant cells accurately. Recently many numbers of image processing and machine learning techniques has been researched and implemented. The primary intention of this study is to analyze the best image segmentation and classifier methods in order to locate the tumour cells accurately. Lung cancer can be detected in three different ways like radiography, MRI (Magnetic resonance Imaging) and

¹Department of Electronics and Communication, Rajalakshmi Institute of Technology, Chennai, India. ²Department of EIE, Saveetha Engineering College, India. ³Department of Medical Electronics, Saveetha Engineering College, Chennai, India. ⁴ECE, Hindusthan Institute of Technology, Coimbatore, India. ⁵Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences Chennai, India. *For Correspondence: malathiyoga08@gmail.com

Computed Tomography (CT). Among all methods, CT is one of the good detection techniques with high accuracy, low cost, short imaging time and widespread availability. Segmentation is the process of separating the region of interest from an acquired CT image. In the recent years many researchers implemented different segmentation techniques. The proposed system uses the aim of this study was to utilize Hybrid segmentation technique in diagnosing tumoral tissue in CT images.

Uzelaltinbulata et al., (2017), has proposed a technique to partition the lung tumour from CT images. Recently, many of the researchers use machine learning to explain the segmentation difficulty. Many image segmentation techniques have more difficulties to find the Region of Interest (ROI). The conventional segmentation methods involve preprocessing to enhance the image qualities and also to reduce the noise. Further with the help of any one of the segmentation techniques, the author segments the tumor portion from the affected lung image.

Kumar et al., (2019) has discuss the significance of optimization algorithms namely k means clustering, Particle Swarm Optimization and Guaranteed Convergence Particle Swarm Optimization (GCP SO) to separate tumour from the acquired CT images. Finally, the author implemented the method called evolutionary algorithms for 20 sample lung images and GCP SO provides the highest accuracy of 95.89%. another study (Subbiahpillai et al., 2020) used segmentation methods in order to find the abnormalities in organs. The author discussed variety of segmentation algorithm but till now there are no universally accepted algorithms for segmentation. The author compares different segmentation algorithms and it has been implemented using MATLAB 2010 software. In another study,

Tanzilasaba (2020), The author implemented a survey to identify the cancer by means of machine learning practices for breast, brain, lung, liver, skin cancer and leukemia. Consequently, the author has discussed the performances of various segmentation characteristics such as sensitivity, specificity, false positive matrices and the author has also listed the challenges for future work.

The CAD (Computer Aided Diagnosis) has been used in order to separate pulmonary nodules (Xiangxia li et al., 2020). But this CAD is a difficult task in the existence of intrinsic noise. The author has implemented FCM to segment Lung cancer CT images, which helps to preserve image details. But in this method spatial information was not considered. Hence this problem can be overcome by Fuzzy C means algorithm. It reduces the computer complexity and helps to improve performance.

GopiKasinathan et al., (2019), proposed computer aided diagnosis to improve a patient's chance of survival. The author has used active contour model for 3D lung segmentation. The recommended technique integrates the local energy term and a multiscale Gaussian distribution helps to identify the tumour from lung images. The researches use lung image database consortium (LIDC- IDRI) data set that has 850 lung- nodule images. The proposed technology provides with the accuracy of 97%. It also uses enhanced CNN classifier to classify the affected region as normal or abnormal.

Malayilshanid and Anitha (2020), implemented an automated lung cancer detection scheme using deep learning and hybrid optimization algorithm. The author has preprocessed the acquired lung image, followed by segmentation of the tumour portion using active contour. Grid based scheme helps to identify the nodules in the segmented image. The author has recommended salp - elephant herding optimization algorithms based on deep belief network (SEOA- DBN) to classify the tumoral tissue. The proposed algorithm provides 96% efficiency.

Computer aided diagnosis (CAD) plays an important role to find the lung cancer in the initial stage. To separate the ROI from the affected portion accurately, preprocessing is the important process (Sahu et al., 2017). This study uses Fuzzy c means with automatic thresholding and morphological processes in order to perform the lung segmentation effectively. The author uses database from LIDCIDRI data set which has 10 normal and abnormal images for segmentation process. The accuracy of the future method was compared with the assistance of reference standard. It has obtained by tracing the lungs region manually with the expert. It provides the 99.94 % of accuracy and also provides 0.94 jaccards index and 0.97 dice similarity co-efficient values.

Jamshidsoltani et al., (2020), has implemented new technology called improved region growing to obtain exact segmentation of pulmonary tumours with a good precision in a shorter time. The algorithm was applied to 4 patient CT images and it gives the accuracy of 98% compared to conventional methods. In another study (Liu, 2020) a novel algorithm to segment the lung tumour in an accurate and automatic way was used. The algorithm uses image decomposition filter to obtain accurate segmentation. The results have provided dice resemblance index of 94.91% on CT lung images when evaluated with ground truth images.

The aim of this study was to segment the tumoral tissue of CT lung cancer image in an effective way.

Materials and Methods

The process flow diagram to perform lung cancer segmentation is represented in Figure 1.

It involves Preprocessing, Hybrid segmentation, feature extraction using GLCM, CNN classifier. The preprocessing involves helps to reduces noises present in the acquired image. Hence the image quality should be improved before the segmentation process. While removing noise and artifact it is necessary to preserve its images without degradation of the original image. The median filter helps to perform the preprocessing operation.

Fuzzy C means algorithm

It is one of the best-unsupervised algorithms used for medical image segmentation. It is developed by DQMM and modified by Bezdek. Fuzzy clustering finds the application in pattern recognition. This is an iterative type clustering method. It is one of the most suitable methods for segmentation when the k value is predefined. A Fuzzy algorithm is capable of processing the overlapped dataset.

Results

The technique is able to give good segmentation results for noise free images. The algorithm is described as follows

Step 1: After assigning the K value, assign the membership value of each data point based on the cluster center and the data point. The foremost intention of the algorithm to minimize the following function

$$J = \sum_{i=1}^N \sum_{j=1}^C \|x_j - c_j\|^2 \quad (1)$$

Where N is the number of data points

C refers the required number of clusters.

δ_{ij} -Defines the value of membership for ith data point x_i in the cluster j

$\|x_j - c_j\|$ -Represents the measure of the closeness of the data point x_i to the center vector c_j of the cluster j

The above formula estimates the distance between the data point and cluster center.

Step 2: Next the data points near to the particular cluster center has the largest membership value of that specific center. The membership value is calculated by using the following formula

Let x_j is the data point have the degree of its membership to that particular cluster j is calculated as

$$\delta_{ij} = \frac{1}{\sum_{k=1}^C \left\| \frac{x_i - C_i}{x_i - C_k} \right\|^{\frac{2}{m-1}}} \quad (2)$$

m-is the fuzziness co-efficient and c_j is calculated as follows

$$C_j = \frac{\sum_{i=1}^N (\delta_{ij}^m x_j)}{\sum_{i=1}^N \delta_{ij}^m} \quad (3)$$

δ_{ij}^m -Degree of membership value.

The membership value lies between 0 to 1. It means $\sum \delta_{ij} = 1$

Step 3: The fuzziness coefficient values lie between $1 < m < \infty$. The values show that the amount of cluster can overlap with each other. Number of iterations of the segmentation process depends on the accuracy of degrees of membership values. This accuracy of degrees of membership is measured using the amount of membership value which actually varies from one iteration to the next.

$$\epsilon = \Delta_i^N \Delta_j^C |\delta_{ij}^{k+1} - \delta_{ij}^k| \quad (4)$$

$\delta_{ij}^{k+1}, \delta_{ij}^k$ are membership values for K, K+1 iteration respectively.

Step 4: The algorithm ends when the cluster center is stabilized.

The disadvantages of FCM are that all the membership values for a data point in the entire cluster are one, but the outlier points have the value more than one. It is difficult to process this kind of phases. The algorithm is more suitable for noise free images. The following are the input lung images after histogram equalization. The various stages of results of Fuzzy C means algorithm shown in Figures (2 ,3 ,4 and 5).

Active Contours

The deformable models are also known as an active contour. Active contour is established by Kass et al in 2-D space, and it can be enhanced for 3 -D space by Terzopoulos et al. An Active contour or snakes are used in 2D space, and balloons are used in 3 D space. Under the influence of an external force, the parametric curve moves inside the image to detect the boundaries of the

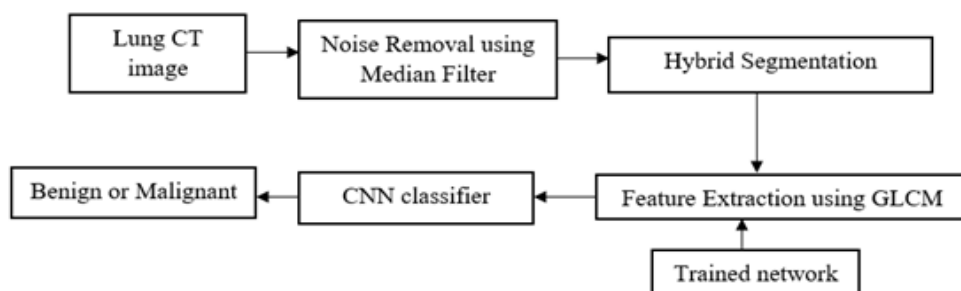


Figure 1. Flow Diagram of Lung Cancer Segmentation

Table 1. Features of Normal and Abnormal Images

Images	Features For Normal Images				Features For Abnormal Images			
	Mean	SD	PSNR	MSE	Mean	SD	PSNR	MSE
IMAGE 1	130.25	52.69	2.78	190	136.68	46.98	77	0.34
IMAGE 2	131.26	50.80	3.90	189	123.46	52.67	50	0.67
IMAGE 3	121.63	45.52	3.44	182	119.23	42.39	78	0.45
IMAGE 4	123.55	56.62	3.89	178	121.73	43.25	89	0.56
IMAGE 5	112.76	46.89	4.01	158	132.25	40.35	91	1.04



Figure 2. Input Image



Figure 4. ROI of Normal and Abnormal Images



Figure 3. Input Image after Histogram Equalization



Figure 5. Fuzzy based Clustering

object. Georges et al., (1999) has discussed regarding deformable model which gives the sequence with a mixture of geometry, physics and approximation theory. The contour of the object is understood from geometrical information; Physics provides information about the way in which the shape changes with more space and time to make the parameter more appropriate to the curve with the help of approximation theory.

Step 1: Place the Active contour or snake subsequently to the region of interest.

Step 2: With the help of internal and external forces created inside the image, the snake is moved close to the object by an iterative process.

Step 3: The energy function is predictable for internal and external forces, which is capable of finding the accuracy of region of interest.

Step 4: The foremost intention of this technique is to diminish the energy function. The internal forces assist to smooth the data set and external forces move forward to the contours close to the region of interest. The results of active contour segmentation is represented in Figures (6 and 7).

Hybrid Segmentation

The conventional algorithm uses the single membership value to characterize the preferred pattern which is not sufficient to perform the segmentation exactly (Chen, 2019). For Lung tumour detection, the pixel intensity is the single parameter that is not enough to classify the brain tissue. When any dissimilar structure appears, the conventional FCM is not sufficient for segmentation.

This can be avoided by adding the spatial information of neighboring pixels which can be considered to define the probability function of each pixel. This spatial information helps to find new membership values for each pixel. It leads to reduce the problem due to noise and intensity in homogeneity and increases the accuracy of the result. Next, we use active contour for segmentation of tumour. It works slowly in case of large image size and it is not capable to segment. A technique called hybrid technique combines both spatial FCM and active contour to overcome the above disadvantages. Active contour method helps to detect the contour of an object by creating a snake around its boundary. Further FCM helps to segment the tumoral tissue automatically from the affected portion.

The results of hybrid segmentation represented as in Figures (8, 9, 10 and 11).

Discussion

Feature Extraction using GLCM

In this paper, gray-level co-occurrence matrix (GLCM), histogram methods and gray-level run length matrix (GLRLM) were utilized to remove features. In general, GLCM were calculated by means of four dissimilar orientations. For each voxel located in the region of interest, the features were removed by taking a patch on that voxel with 7x7x7 neighborhood. The extracted features are listed in the following Table 1.

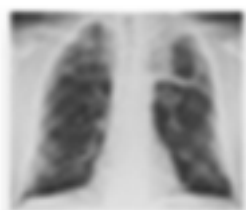


Figure 6. Original Image

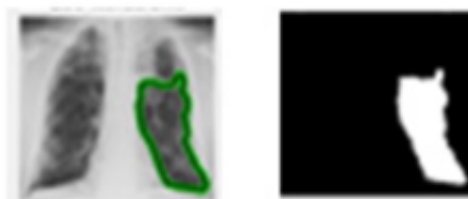


Figure 7. Segmentation Using Active Contour

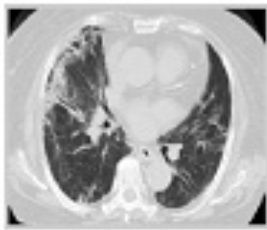


Figure 8. Input CT Image



Figure 10. Classified Image Using Existing Method

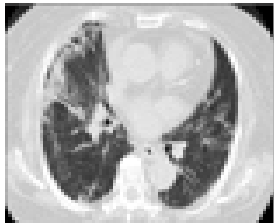


Figure 9. Gaussian Scale Space Filtered Image



Figure 11. Hybrid Segmentation Image

Convolutional Neural Network

The architecture of CNN is depicted in Figure 11. The network helps to categorize the nodules. The network has three layers namely, input layer, hidden layer, and output layer. There are $P \times P$ neurons in the input layer that represent the $P \times P$ pixel of the image obtained from segmentation process Yu et al., (2020). The hidden layer has n groups of $N \times N$ neurons structured as over $N \times N$ feature map (where $N = P - r + 1$) and the $r \times r$ area is represented as the interested area. Hidden neuron selects input from $r \times r$ neighboring piece in the input image section. If the neurons with the similar feature map are one neuron distant, then their interested areas in the input layer are one pixel distant shown in Figure 12. Each neuron of the related feature map is kept to obtain the same group of r^2 weights and achieve the equal action on the resultant fragments of the input image.

The benefit of hindering the weights permits the network to accomplish pattern recognition, which is

shift-invariant. Hence, the entire action is represented as $r \times r$ convolution kernel. The feature map output is achieved from the convolution of the input with the $r \times r$ convolution kernel. Every hidden neuron y_j creates its output by means of an activation function represented as in (5). The maximum and minimum activation functions are correspondingly zero and one.

$$y_j(w, x, a_j) = f\left(\sum_{i=1}^{r^2} (w_{ji}x_i) + a_j\right) \quad (5)$$

w_{ji} - the weight between the pixel and hidden neuron, j, i of the input image.

x_i - gray value of the input pixel i .

a_j - the bias of the hidden neuron j .

x_1, x_2, \dots, x_{r^2} - Input image pixels, each connected with every neuron j .

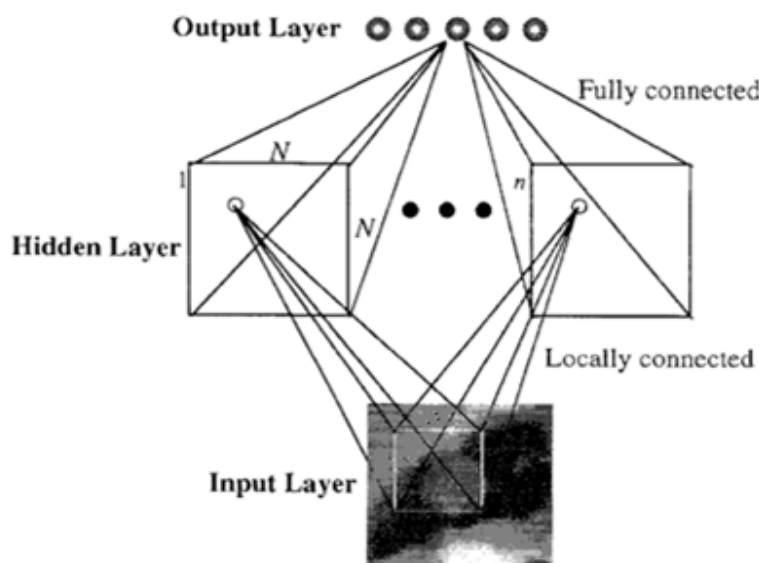


Figure 12. Architecture of One Hidden Layer CNN

The output layer is completely associated with the hidden layer. The sigmoid activation function, z_0 of the output neuron is characterized by,

$$z_0(w, y, g_0) = \frac{1}{1 + \exp\left\{-\left[\sum_{j=1}^{nN^2} (w_{0j}y_j) + g_0\right]\right\}} \quad (6)$$

w_{0j} -Weight between the neuron and the output neuron in the hidden layer j
 nN^2 - total number of neurons in the hidden layer
 g_0 - bias of the output neuron.

Hence, the system contains $(O+P^2+nN^2)$ number of neurons and $(nN^2 (r^2+O+ 1) +O)$ number of links. These numbers include the input neurons and bias links also. The number of independent links is given by $nN^2 (O+ 1) +nk^2+O$. O represents the count of output neurons.

In Rumelhart et al., (1985), both network weights and the bias weights are distorted by the application of Back Propagation (BP) algorithm. The BP algorithm iteratively alters the weights with the intention of reducing the whole error of the actual output vector from the target vector. The error function to be reduced is called as the Sum-of-Squared Error (SSE). During training, the concerned areas surrounded by one hidden class are limited to consume the equal form of weights. The weights between output and hidden layers and every involved area weight, are altered by means of stochastic mode. In this method, for every training sample the weight difference is acquired from back-propagated error and is changed directly for every neuron.

In conclusion, the segmentation and classification of medical images is important in diagnosing several diseases. Hence the proposed work uses hybrid technique which means the combination of active contour and spatial fuzzy C means clustering algorithm. Finally, the results of the implemented technique provided good accuracy, Peak signal to noise ratio (PSNR), Mean Square Error (MSE) value. In future the other techniques can be utilized to improve the details before segmentation. The proposed work provides 96.67 % accuracy.

Author Contribution Statement

All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

None.

Conflict of Interest

There is no conflict of interest.

References

Caixia L, Minhyong P (2020). Automatic lung segmentation based on image decomposition and wavelet transform. *Biomed Signal Process Control*, **61**, 1-9
 Georges BA, Jyrki N, Nancy W, Benoit MD (1999). Model

Creation and Deformation for the Automatic Segmentation of the Brain in MR Images. *IEEE Trans Biomed Eng*, **46**, 1346-56.

Gopi K, Selvakumar J, Amir HG, et al (2019). Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier. *Exp Sys Appl*, **134**, 112-9.

Heng Y, Zhiqing Z, Qiming W (2020). Deep learning assisted predict of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model. *IEEE Access*, **8**, 86400-10.

Jamshidsoltani N, Abdollah K, Behrooz N (2020). Lung tumour segmentation using improved region growing algorithm. *Nucl Engin Technol*, **52**, 2313-9.

Malayilshaid A (2020). Lung cancer detection from CT images using salp-elephant optimization-based deep learning. *Biomed Engin Appl Basics Commun*, **32**, 2050001- 14.

Rumel hart DE, Hinton GE, Williams RJ (1985). Learning internal representations by error propagation (No. ICS-8506)', California Univ San Diego La Jolla Inst for Cognitive Science.

Satyaprakash S, Priyanka A, Narendra DL, Shrish V (2017). A new hybrid approach using Fuzzy clustering and morphological operations for lung segmentation in thoracic CT images. *Biomed Pharmacol J*, **10**, 1949-61.

Selin U, Buse U (2017). Lung tumor segmentation algorithm, 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 24-25 August, Budapest, Hungary, Procedia Computer Science 120, pp 140 –7.

Senthil Kumar K, Venkata Lakshmi K, Karthikeyan K (2019). Lung cancer detection using Image segmentation by means of various evolutionary algorithms. *Comput Math Methods Med*, **2019**, 1-16.

Subbiahpillai NK, Alfred LF, Paul SV (2020). An overview of segmentation algorithms for the analysis of anomalies on Medical Images, Intelligent Systems. *J Infect Public Health*, **29**, 612–25.

Tanzilasaba (2020). Recent Advancement in cancer detection using machine learning: systematic survey of decades, comparison and challenges. *J Infect Public Health*, **13**, 1274-89.

Wei C, Haifeng W, Suting P, et al (2019). Hybrid segmentation network for small cell lung cancer segmentation. *IEEE Access*, **7**, 75591-602.

Xiangxia Li, Bin Li, Fang L, Hua Y, Fengzhou (2020). Segmentation of pulmonary nodules using a GMM Fuzzy C-means algorithm. *IEEE Access Engin Med Biol Soc Section*, **8**, 37541 – 56.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.