

Automatic Lung Segmentation and Lung Nodule Type Identification over LIDC-IDRI dataset

R. Jenkin Suji

WOSA Researcher, ABV-IIITM Gwalior,
Gwalior, Madhya Pradesh – 474015, India
sujjenkin@gmail.com

Sarita Singh Bhadauria

Professor, RGPV Bhopal
Bhopal, Madhya Pradesh – 462033, India
saritamits61@yahoo.co.in

Abstract

Accurate segmentation of lung parenchyma is one of the basic steps for lung nodule detection and diagnosis. Using thresholding and morphology based methods for lung parenchyma segmentation is challenging due to the homogeneous intensities present in lung images. Further, typically, datasets do not contain explicit labels of their nodule types and there little literature on how to typify nodules into different nodule types eventhough identifying nodule types help to understand and explain the progress and shortcomings of various steps in the computer aided diagnosis pipeline. Hence, this work also presents methods for identification of nodule types, juxta-vascular, juxta-pleural and isolated. This work presents thresholding and morphological operation based methods for both lung segmentation and lung nodule type identification. Thresholding and morphology based methods have been chosen over sophisticated approaches due to the reasons of simplicity and rapidity. Qualitative validation of the proposed lung segmentation method is provided in terms of step by step output on a scan from LIDC-IDRI dataset and lung nodule type identification method is provided by output volume images. Further, the lung segmentation method is validated by percentage of overlap and the results on nodule type identification for various lung segmentation outputs have been analysed. The provided analysis offers a peekview into the ability to analyse the lung segmentation algorithms and nodule detection and segmentation algorithms interms of nodule types and motivates the need to provide nodule type groundtruth information also for developing better nodule type classification/identification algorithms.

Keywords: Lung Segmentation; Juxta-vascular nodules; Juxta-pleural nodules; Thresholding; Morphological operations.

1. Introduction

Lung cancer [6] is one of the leading cause of cancer deaths in the world. Manual diagnosis of lung cancer is prone to inter reader and intra reader variations. Pulmonary nodules are the most important symptom for lung cancer. Automatic CAD system for lung cancer is needed for quick response and early detection of lung nodules. The steps in CAD system for lung cancer detection and diagnosis are preprocessing, lung segmentation, lung nodule candidate detection, lung nodule segmentation, feature extraction, lung nodule classification and false positive reduction.

Lung Segmentation is one of the major steps in nodule detection and accurate segmentation of the lung structures is vital for an automated CAD system for Lung Cancer since it reduces the search space for detecting the lung nodules while also improving the diagnosis. But the challenges for lung segmentation are the homogeneities that exist in lung structures (i.e) lungs share similar intensities to the neighboring structures of the lung.

The publicly available lung segmentation datasets are LObe and Lung Analysis 2011 [14] and Lung CT Segmentation Challenge 2017 [15]. LIDC-IDRI dataset [3] comprises diagnostic lung cancer scans with marked up annotated lesions. It is for development and evaluation of computer assisted diagnostic (CAD) methods for lung cancer detection and diagnosis.

Based on the location and attachment of nodules, pulmonary nodules are classified into well circumscribed/isolated nodules, juxta-vascular nodules, juxta-pleural nodules and nodules with pleural tail [6]. Isolated nodules are nodules which is not connected to the neighboring structures of the lung, juxta-vascular nodules are nodules attached to the blood vessels of the lung, juxta-pleural nodules are nodules attached to pleural

wall of the lung and the nodules with pleural tail have tail like structure but not attached to the pleural surface of the lung. Similarly, based on texture, pulmonary nodules are classified into solid nodules, part solid nodules and non solid/ GGO (Ground Glass nodules. Solid nodules are classified when texture = 4 or 5, part solid nodules are classified when texture = 2 or 3 and when texture = 1, it is called as non solid/ GGO nodules [18]. According to size, pulmonary nodules are classified into small and large nodules [22].

Lung nodule detection plays a major role in the detection and diagnosis of lung cancer. Lung nodule detection is challenging task and one way to reduce the complexity of finding nodules is to reduce the search space. This is performed by segmenting the lung lobes before performing candidate nodule detection. This paper discusses two simple nodule detection methods and evaluates them based on ground truth overlap measure in the absence of ground truth lung segmentation in the LIDC-IDRI dataset.

Further, LIDC-IDRI dataset contains different types of nodules such as juxta-pleural, juxta-vascular, ground glass opacity and isolated nodules etc. But the identification of the nodules from the dataset in these lines is absent except for ground glass opacity nodules (nodules with texture =1 are normally identified as ground glass opacity nodules [18]. This paper discusses two methodology for identifying a nodule as juxta-vascular and juxta-pleural. The nodules which are neither juxta-pleural and juxta-vascular are identified as isolated.

The rest of the paper is organized as follows. Section.2 gives the literature survey of the various lung segmentation methods. Section.3 describes the step by step approach for lung parenchyma segmentation and approach of identifying the juxta-vascular, juxta-pleural and isolated nodules from the LIDC-IDRI dataset. Section.4 gives the results and discussions by calculating the percentage of overlap and Section.5 concludes the discussion with conclusion and future direction.

2. Literature Survey

This section provides a survey on lung parenchyma segmentation and lung nodule type identification.

2.1. Lung Parenchyma Segmentation

Traditional lung parenchyma segmentation methods may be classified as thresholding based, region growing based, model based or template based, neighboring anatomy based and machine learning based [16].

Thresholding [2,23] methods are fast and least expensive, and work by selecting a particular threshold which separates the lung structures from other body parts. Using a same threshold to segment the lungs is not appropriate since due to variations in image acquisition equipments and process image contrast intensities may vary. Optimal thresholding [5] then selects the threshold automatically according to the gray values of the pixel neighborhood, but it fails to include the nodules when the nodules are attached to the pleural surface or vessel since it separates the nodules as fat [12]. Morphological operations along with optimal thresholding are performed to include the attached nodules [11].

Region growing methods [1] work by selecting a manual/automatic seed point to separate the high density areas of the lung such as ribs, tissues etc which are attached to the lung parenchyma based on the homogeneity criteria. On the other hand, region growing methods [24] may fail to separate the lung lobes and since it depends on accurate seed points which may affect the stability. Model/Atlas based methods [13] use prior shape for segmenting the medical images and this leads to inaccurate segmentation if the model is not close to the actual boundary. Machine learning [12,24] methods require a large training data for extracting the features. Below, there is a survey on literature.

Pramod and Mrityunjaya [12] implemented a fully automatic lung parenchyma segmentation using 2D optimal thresholding for binarization, 2D reconstruction to remove neighbouring anatomical structures, improved lung chain code to repair the lung boundaries and Bresenham pixel reconstruction for including the juxta-pleural nodules. They evaluated their work-flow using 115 thoracic CT lung parenchyma scans which includes 21 slices having juxta-pleural nodules of LIDC-IDRI, lung TIME1 and lung TIME2 datasets with privately annotated ground truth and achieved greater than 96% accuracy for lung parenchyma segmentation and 100% sensitivity which includes nodules attached to the pleural wall. Mesanovic et al. [17] proposed a method based on region growing and the edges are removed using edge operations along with morphological operations. They used 50 scans of slice thickness of 1 to 5 mm from privately annotated dataset and they missed the nodules attached to the pleural wall. Some researchers used rolling ball method to include the juxta-pleural nodules but they found it difficult to accurately set the size of the sphere [2].

In order to include the juxta-pleural nodules attached to the pleural wall, Wei et al. [24] implemented a fully automatic segmentation method and repairing strategy based on optimal iterative thresholding, 3D connected component labeling, 3D region growing, center position locating method, improved chain code and Bresenham algorithm. They used 97 thoracic CT scans and 25 juxta-pleural nodules from Lung Nodule Database and the Juxta-pleural Nodule Database from ShengJing Hospital and they proved the 100% sensitivity, 98.6% accuracy for including juxta-pleural nodules and the accuracy of lung parenchyma segmentation is >95.2%. Their approach took only 0.67 sec per scan for parenchyma segmentation. As the CT scans contain difference in image data, noise received from the detectors or patients during procedures by technician, Zhang et al. [26] used

otsu thresholding and improved grab cut for lung parenchyma segmentation which is insensitive to noise. Preprocessing methods like median filtering, wiener filtering, smoothing filters etc were used to remove the noise in the CT images [11]. Xiao et al. [25] used iterated threshold along with fractal geometry and improved convex hull for including juxta-pleural nodules in the lung parenchyma region. They used 97 cases from the private hospital and obtained 92.45% pixel accuracy and 95.9% intersection over union.

Some of the researchers performed lung parenchyma segmentation as a part of their lung nodule detection methods and they used lung nodule detection datasets. Gong et al. [8] used Otsu thresholding, 3D region growing, airway removal and boundary repairing for lung parenchyma segmentation. They also used 3D tensor filtering with local feature analysis to refine the attached lung boundaries. Zhang et al. [27] used global optimal active contour model for lung parenchyma segmentation and used voxel remove rate and 3D features by characterizing the attached nodules. Naqi et al. [19] used corner seeded region growing along with differential evolution based optimal thresholding, morphological operations, hole filling and juxta-vascular nodule extraction for extracting the lung region. Gupta et al.[9] proposed thresholding along with morphological closing for lung parenchyma extraction.

Setio et al. [22] proposed a method for lung region extraction based on large airway extraction, lung segmentation, left and right lungs separation, segmentation smoothing and segmentation refinement to include the pleural nodules. Choi and Choi [5] used optimal thresholding, 3D Connected Component labeling, lung contour refinement for lung paraenichyma extraction.

Cha et. al [4], Farhangi *et al.* [7] and Netto et al. [20] have used other methods such as graph cut, adaptive shape prior, motion information, 3D Active Contour Segmentation based on Sparse Linear Combination of Training Shapes, volume of interest reduction etc. for lung nodule detection. The above methods mostly used datasets specific to lung segmentation, private datasets or data from hospitals with manually annotated ground truth by radiologists. Hence the lung segmentation efficiency was calculated by comparing the segmented image with the ground truth in terms of sensitivity, specificity, accuracy, precision etc. Since the LIDC-IDRI dataset has only annotated nodule regions, a percentage overlap measure was used for checking the efficiency of the lung segmentation by comparing the segmented region and the ground truth of the nodule region as done in Setio et al [22].

2.2. Lung Nodule Type Identification

Literature is brim with research work on segmenting juxta-vascular and juxta-pleural nodules from their attachments. But, a procedure to automatically identify a nodule type as juxta-pleural or juxta-vascular is absent in the literature. This section presents a couple of procedures for identifying the type of a nodule as juxta-pleural or juxta-vascular and compares the same.

The major challenges to performing lung nodule type identification are the following:

1. Distinguishing vessels from nodules or vessels by intensity alone is non-trivial due to the homogeneity of these regions since the CT values of the nodules as well as these non-nodules are very similar.
2. Due to the above reason, it is difficult to differentiate between the nodule and lung boundaries as well as the nodule and vasculature boundaries.

The contributions of the proposed work are

1. Thresholding and morphological operations based method for lung parenchyma segmentation on scans from LIDC-IDRI dataset.
2. Methodology for identifying juxta-vascular, juxta-pleural and isolated type nodules in LIDC-IDRI dataset.
3. Evaluation and comparative analysis of the results obtained

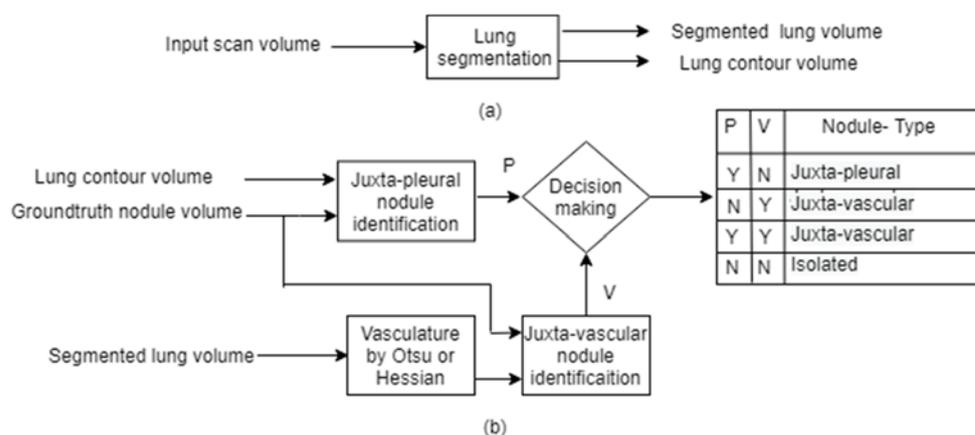


Fig. 1. Proposed Methodology

3. Proposed Methodology

This section explains the methodology for contributions 1 & 2 listed above. Figure 1(a) explains the overall procedure of the work flow for lung segmentation as described in section. Figure 1(b) describes the overall workflow for identifying a nodules as Juxta-Pleural or Juxta-Vascular through the key methods as described in this section.

3.1. Lung Segmentation

Lung segmentation involves segmenting the lung which means to delineate (or) separate the lung from the surrounding organs and structures such as the heart, rib cage, and diaphragm. The segmented lung comprises of parenchyma, which is a cellular tissue lying between the lung body wall and the organs. This is present in two lobes – right lobe and left lobe. The terms "lung segmentation" and "lung parenchyma segmentation" have been used synonymously in this paper.

Lung Parenchyma segmentation is one of the important step for further detecting the lung nodules accurately. One of the simplest method for segmenting the lung region is thresholding. Fixed thresholding uses constant threshold value on intensity range for thresholding all the scans and most of the nodules from the groundtruth stay outside the segmented parenchymal region. On the other hand Otsu optimal thresholding adaptively selects the threshold based on the neighbourhood region. The procedure adopted in this work is elaborated as below:

3.1.1. Lung Parenchyma Segmentation

The step by step approach for lung parenchyma segmentation given below is applied on every slice of a given scan volume. The following procedure has been applied on LIDC-IDRI-0067, slice number 113 as shown in figure 2(a) and the results are shown through the following steps.

Input : Lung Scan Volume

1. Identify the minimum HU voxel intensity value from the volume and make it to zero to obtain the image shown in figure 2(b).
2. Apply Otsu thresholding to obtain the image as displayed in figure 2(c). It is to be noted that fixed thresholding will not be uniformly applicable on all scans due to the reason that different scans have varying contrast and resolution based on different capture devices.
3. Find the number of connected components from the thresholded binary image as well as the negated thresholded binary image. The negated thresholded image is shown in figure 2(d).
4. Check if number of connected components in step 3 are greater than 100 for both thresholded binary image and its negative binary image and if so, apply median filtering using a disk structuring element of radius 5, to remove the effect of salt and pepper noise from the image. The resultant image is shown in figure 2(e). The value 100 for thresholding the number of connected components has been selected arbitrarily after few trials with other values.
5. Use a disk structuring element of radius 5 and apply erosion to remove the pleural attachments where the nodules are attached to the wall. It also helps in separating the lung nodules attached to blood vessels. Now, pad the resultant image with zeros on the entire edge of the image on four sides as displayed in figure 2(f) in-order to allow for using the next step for proper segregation of the region outside the lung body from the body of the lung. The image thus processed is going to contain one large component which is the spine and the body region that surrounds the lung region. Now store this large component region by finding the number of connected components and selecting the first component and place only this large component region on a zero-initialized matrix of size 512×512 as shown in figure 2(g).
6. Now negate this image and find the number of the connected components.
7. Negate the first component. This gives the inverted segmented parenchyma mask as shown in figure 2(h). The negation of this gives the segmentation mask as shown in figure 2(i). The contour of this mask for the entire volume obtained using roberts edge detector for each slice can be stored as the parenchyma volume for use in juxta-pleural nodule identification process.
8. This mask can be then used to segment the Lung Lobe from the original image as given in figure 2(j).

Lung Segmentation efficiency is evaluated based on percentage of overlap between the segmented lung region with the ground truth of the nodule region. This efficacy of this segmentation approach is discussed in section.4.

3.2. Lung Nodule Type Identification

Lung nodules are classified based on location as juxta-vascular nodules, juxta-pleural nodules and isolated nodules, based on texture as solid nodules, part-solid nodules and non- solid/ Ground-Glass Opacity (GGO) nodules. Based on the diameter of the nodules, they are classified as small nodules or large nodules. This subsection describes the procedures for checking if a nodule is juxta-pleural and juxta-vascular.

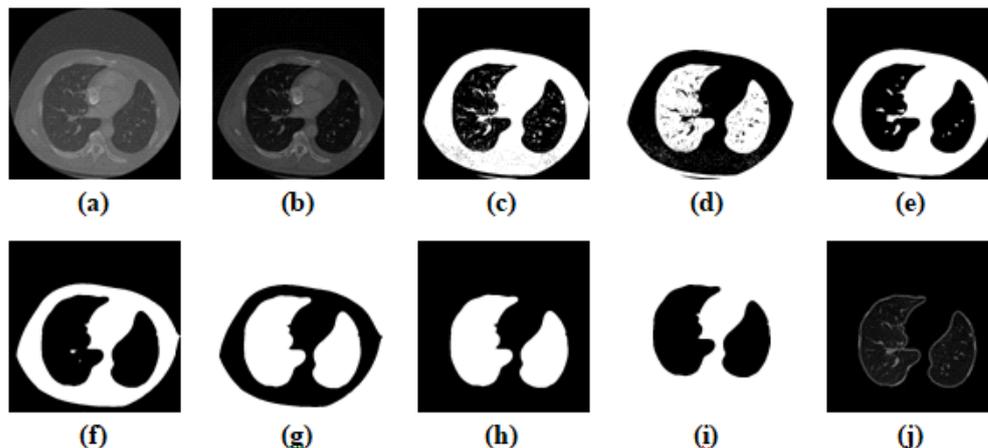


Fig. 2. (a) original image (b) min replaced image (c) otsu thresholded image (d) negated otsu image (e) median filter applied (f) eroded and padded image and largest component computed (g) largest component (body) plotted on a plane square (h) negation of the largest component (i) invert the previous image to obtain lung mask (j) lung segmented image using lung mask

3.2.1. Juxta-pleural Nodule Identification

The procedure for juxta-pleural nodule identification is given below by applying the procedure on LIDC-IDRI-0001.

Input : Lung Scan Volume

1. Extract the Lung Contour from the lung scan volume : Use the procedure (step 1 to 7) discussed in Lung Parenchyma segmentation to obtain the lung contour for the entire lung scan volume. The resultant contour volume is shown in figure 3(a)

2. Identify the nodule from the lung scan volume as lung nodule volume: Use any of the nodule segmentation procedure from the literature [18] to segment the nodule or use the groundtruth nodule segmentation available with standard dataset such as LIDC-IDRI to obtain the nodule groundtruth volume or lung nodule volume. The nodule groundtruth volume for nodule 0 for LIDC-IDRI-0001 is shown in figure 3(b).

3. Check if there is overlap between the lung external contour and the lung nodule volume: Apply dilation operation on the nodule groundtruth volume with a ball structuring element of radius 1 to get dilated lung nodule volume. This dilation operation is necessiated as a measure to overcompensate for any missing nodule voxels that may not have been segmented as the nodule. Figure 3(c) shows the overlap that occurs between the images (figure3(a), figure3(b)) and is highlighted by circle.

4. The degree of overlap determines the degree of juxta-pleurality of the nodule. This degree of overlap is calculated by performing a "bitwise-and" operation between the lung parenchyma volume and the dilated lung nodule volume as shown in equation (1) below:

$$\text{Degree of overlap} = \text{Bitwise - And (Lung Nodule Volume, Lung Contour Volume)} \dots\dots\dots(1)$$

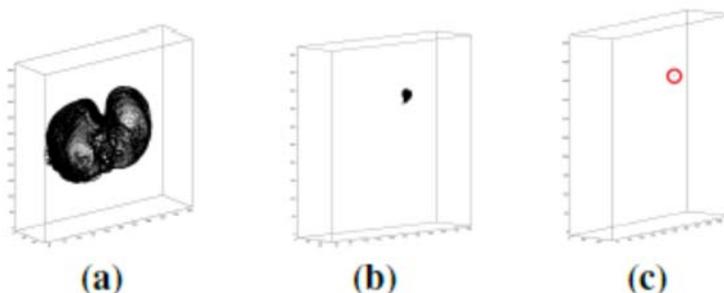


Fig. 3. Otsu based Juxta-pleural nodule type Identification (a) Lung Contour Volume (b) Lung Nodule Volume (c) Overlap between Lung Nodule Volume and Lung Contour Volume

3.2.2. Juxta-vascular Nodule Identification

The procedure for juxta-vascular nodule identification is given below by applying the procedure on LIDC-IDRI-0007.

Input : Lung Scan Volume

1. Use the procedure (steps 1 to 8) discussed in Lung Parenchyma segmentation to obtain the segmented lung volume. This volume is eroded with a ball structuring element of radius 5 to get eroded lung segmented volume. This operation is necessiated to remove the outer lining of the lung.

2. Extract the lung vessel volume from the volume obtained after step 1. This is obtained either using otsu thresholding as displayed in figure 4(a) or Hessian based approach as shown in figure 5(a) : In otsu thresholding approach, the otsu thresholding is applied on segmented and eroded lung volume to obtain the lung vessel volume. This volume typically also contains a nodule. In the hessian approach, hessian recursive gaussian image filter (from ITK library) is applied over the volume from previous step and a threshold of 150 is used to obtain the lung vessel volume.

3. Use any of the nodule segmentation procedure from the literature [18] to segment the nodule or use the groundtruth nodule segmentation available with standard dataset such as LIDC-IDRI to obtain the nodule groundtruth volume or lung nodule volume. The groundtruth volume for LIDC-IDRI-0007, nodule 0 is given in figure 4(b) and figure 5(b). Both these figures are the same.

4. Apply dilation operation on the nodule groundtruth volume with a ball structuring element of radius 2 to get dilated lung nodule volume: This dilation operation is necessiated as a measure to overcompensate for any missing nodule voxels that may not have been segmented as the nodule.

5. Now, since the lung vessel volume obtained in step 2 may contain nodule also, this requires to be removed from the lung vessel volume shown in figure 4(c) and figure 5(c). Hence remove the dilated lung nodule volume, obtained from previous step, from the lung vessel volume to get nodule removed lung vessel volume.

6. Once more apply dilation operation on the nodule groundtruth volume with a ball structuring element of radius 3 to get dilated lung nodule volume.

7. The degree of overlap determines the degree of juxta-vascularity of the nodule. This degree of overlap is calculated by performing a "bitwise-and" operation, as given in equation (2) between the dilated lung nodule volume and the nodule removed lung vessel volume and the results obtained in case of Otsu based and Hessian based vessel segmentation approaches are shown in figure 4(d) and figure 5(d) respectively.

$$Degree\ of\ overlap = Bitwise - And (Lung\ Nodule\ Volume, Lung\ Vessel\ Volume) \dots\dots\dots(2)$$

The overall process for lung nodule type identification is given in figure 1. Lung Segmentation algorithm takes a lung scan volume as input and gives out a segmented lung volume as well as lung contour volume. This is shown in figure 1(a). For identification of juxta-pleural nodule, the procedure is input with lung contour volume and ground truth nodule volume which output if a nodule is possibly juxta-pleural or not (P=Y or N). For identification of juxta-vascular nodule, first the blood vessel volume is obtained using otsu or hessian from segmented lung volume. Now, the groundtruth nodule volume and the lung vessel volume is input to procedure for juxta-vascular nodules identification which possibly identifies a nodule as juxta-vascular or not (V=Y or N). These two outputs are then fed through a decision making block where the final decision on if a nodule is juxta-vascular or juxta-pleural or isolated is made as following:

$$Nodule - Type = \begin{cases} Juxta - Pleural, if (P = Y) and (V = N) \\ Juxta - Vascular, if (P = N) and (V = Y) \\ Juxta - Vascular, if (P = Y) and (V = Y) \\ Isolated, if (P = N) and (V = N) \end{cases} \dots\dots\dots(3)$$

It is decided to identify a nodule as juxta-vascular if P=Y and V=Y, since in that case the nodule may be appearing along the inner anterior wall of the lobes from where vessels originate.

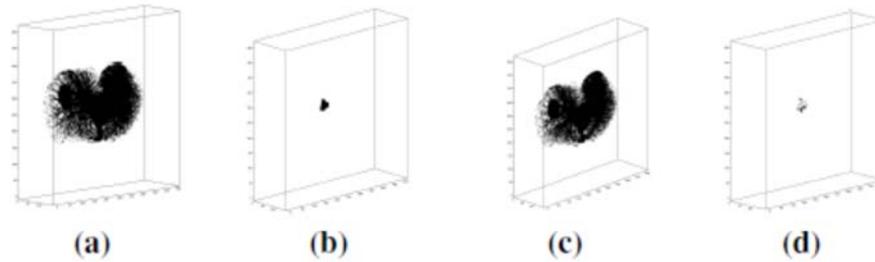


Fig 4 : Lung volume outputs obtained by applying Otsu based vessel segmentation approach for Juxta-vascular nodule type Identification (a) Lung Vessel Volume with nodule (b) Lung Nodule Volume (c) Lung Vessel Volume with nodule removed (d) Overlap between Lung Vessel Volume and Lung Nodule Volume

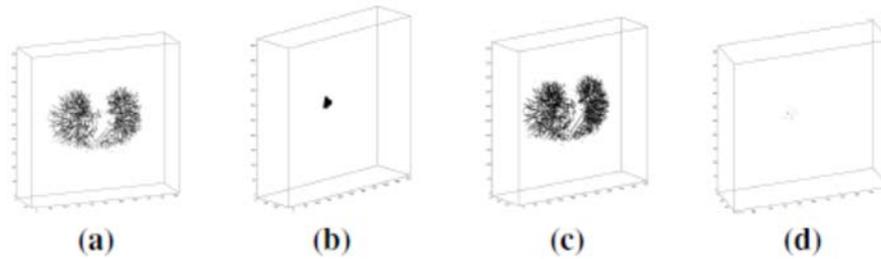


Fig 5 Lung volume outputs obtained by applying Hessian based vessel segmentation approach for Juxta-vascular nodule type Identification (a) Lung Vessel Volume with nodule (b) Lung Nodule Volume (c) Lung Vessel Volume with nodule removed (d) Overlap between Lung Vessel Volume and Lung Nodule Volume

4. Results and Discussions

4.1. LIDC-IDRI Dataset

The Lung Image Database Consortium and Image Database Resource Initiative (LIDC- IDRI) dataset consists of 1010 patient with 1018 thoracic and diagnostic CT scans. Each scan folder consists of multiple 2D slices and a corresponding xml file which contains annotations by four experienced radiologists in blinded and unblinded manner to render an accurate decision. The slice thickness of each scan varies from 1 to 5 mm and pixel spacing varies from 0.6 to 0.8 mm. The dimension of each slice is 512 x 512 pixels.

The step by step procedure for the lung parenchyma segmentation as described in section. 3.1 using Otsu thresholding and morphological operations was applied on the entire LIDC-IDRI dataset.

The LIDC-IDRI dataset has 1018 scans. The total number of nodules on LIDC-IDRI dataset are 2669. 21 nodules were omitted due to mismatch in annotations and 2648 nodules and corresponding annotations are considered for further evaluation. Pylidc [10] a python library, was used for programming and deriving the image results.

S.No	Lung Segmentation Method	Average Percentage Overlap
1	Constant Threshold - 604 [21]	53%
2	Otsu Optimal Thresholding	57%
3	Otsu Optimal Thresholding with assigning minimum = 0	94.2%
4	Otsu Optimal Thresholding with Median Filtering with labels >100	97.4%
5	Otsu Optimal Thresholding Median Filtering with labels >100 and neglabels >100	97.8%

Table 1. Comparison of Thresholding and Morphological operation based methods for lungsegmentation by Average Percentage Overlap

S. No	Slice Thickness	Total number of nodules	Number of nodules with 100 % overlap	Percentage (%) of nodules with 100% overlap
1	0.6	11	11	100
2	0.775	112	100	89.29
3	0.9	6	5	83.33
4	1	195	172	88.21
5	1.25	888	785	88.4
6	1.5	15	9	60
7	2	388	340	87.62
8	2.5	674	592	87.83
9	3	344	287	83.43
10	4	7	4	57.14
11	5	8	7	87.15
Total Number of Nodules		2648	2312	87.31

Table 2 Number of nodules in volumes of varying slice thickness and corresponding number of nodules which are 100% within lung segmentation obtained by proposed method

Table.1 describes the average percentage of overlap of various lung segmentation methods on LIDC-IDRI dataset.

In order to evaluate these different lung segmentation methods, average percentage overlap measure (4) was used.

$$\text{Average percentage of overlap} = \frac{\text{Number of nodule pixels in lung segmentation area}}{\text{total number of pixels in nodule volume}} \dots\dots\dots(4)$$

In method 1, a constant threshold of 604 with the steps as described [21] were used for lung segmentation. In method 2, otsu thresholding was used instead of constant thresholding and rest of the steps were as in [21]. In method 3, the process described in section 3 was used without step 4 (median filtering). In method 4, median filtering was applied only when number of components in the thresholded image is greater than 100. In method 5 the thresholded image as well its negative were checked to apply median filtering. As can be noted here, these steps resulted in an improvement of average percentage overlap to 97.8%. Though the average percentage of overlap of nodules is 97.8%, it is also to be noted that the percentage overlap of voxels is only 87.88%. This means huge chunk of voxels in large nodules are missed through this lung segmentation approach than smaller nodules.

Table. 2 shows the total number of nodules in volumes of varying thickness in LIDC- IDRI dataset and the number of nodules in each category which fully belong within the lung segmentation obtained using the proposed lung segmentation approach. The relation between slice thickness and the ability of the proposed method to successfully segment nodules within lung region is difficult to establish from the percentage overlap column.

S.No	No of Annotations	Total numberof nodules	Number of nodules with 100 % overlap	Percentage (%)
1	1	772	698	90.41
2	2	483	426	88.20
3	3	484	435	89.88
4	4	896	742	82.81
5	>5	13	11	84.62
Total Number of Nodules		2648	2312	87.31

Table 3. Number of nodules with varying number of annotations and corresponding number of nodules which are 100% within lung segmentation obtained by proposed method

Table. 3 shows the number of nodules with varying annotations and those that have been well segmented by proposed method. This signifies the number of nodules with lesser annotations are well within the lung segmentation region than those with large number of annotations.

Table. 4 shows the classification of lung nodules based on Nodule Diameter and Texture as given in the annotated xml file. Among large nodules (>10 mm, with 4 annotations, the number of solid nodules are 421 and the number of part solid nodules are 26 and number of non solid (GGO - round Glass Opacity) nodules are 17.

SN	Radiologist Annotations (atleast)	Nodule Diameter $\geq 3\text{mm}$	Nodule Dia $\geq 10\text{mm}$	Dia $\geq 10\text{mm}^*$	Dia $\geq 10\text{mm}^{**}$	Dia $\geq 10\text{mm}^{***}$
1	1	2643	824	660	79	85
2	2	1874	704	588	64	52
3	3	1393	595	522	47	26
4	4	909	464	421	26	17
	Total	6819	2587	2191	216	180
			Dia. - Diameter	* - Solid	** - Part Solid	*** - Non Solid

Table 4. Number of lung nodules with varying number of annotations and different NoduleDiameters and Textures nodules within lung region is difficult to establish from the percentage overlap column.

S N.	Total Numberof Nodules	Otsu based vasculature			Hessian based vasculature		
		JP	JV	I	JP	JV	I
1	2648	70	1836	742	589	1900	159
2	336	36	298	2	241	9	86

Table 5. Number of nodules which have been identified as juxta-vascular nodule, juxta-pleural nodules and isolated nodules on LIDC-IDRI dataset with 100% Overlap and number of nodules in these types which do not have 100% overlap (JP – Juxta-Pleural, JV-Juxtas-Vascular,I - Isolated)

SN	Type of Nodule	Number of nodules	Otsu based vasculature			Hessian based vasculature		
			JP	JV	I	JP	JV	I
1	Solid nodules	421	5	391	25	180	209	32
2	Part solid nodules	26	0	24	2	14	10	2
2	Non solid/ GGO nodules	17	0	16	1	6	11	0

Table 6. Number of nodules classified with varying texture and nodule diameter $\geq 10\text{mm}$ with greater than or equal to four annotations using proposed approaches, overlap (JP – Juxta-Pleural, JV-Juxtas-Vascular,I - Isolated)

The first row of Table. 5 shows the total number of juxta-pleural, juxta-vascular and isolated nodules as found by the proposed methods. The exact list of nodules which are found to be juxta-pleural, juxta-vascular and isolated is available at http://github.com/sujijenkin/lung_nodule_type_identification. Such list or ground truth is so far not available anywhere to be able to verify if a proposed lung nodule detection approach is well suited for a specific kind of nodule and this may provide a baseline for developing techniques specialized for specific kind of nodules. The second row of Table.5 shows the number of nodules which were not 100% within the lung segmented by the proposed approach. It can also be noted that large portion of nodules which are missed are either juxta- pleural or juxta-vascular. Hence, with such information, better analysis can be performed. Table. 6 shows the texturewise distribution of nodules from LIDC-IDRI dataset with atleast 4 annotations and greater than 10 mm (large nodules) interms of juxta-vascular nodule, juxta-pleural nodule and isolated nodule types. In type identification based on otsu based vessel segmentation, juxta-pleural nodule types are very few in number while in hessian based vessel segmentation, the nodule types are more or less well distributed between juxta-pleural and juxta-vascular nodule types.

5. Conclusion and Future work

The major contributions of this work are two fold : firstly, a thresholding and morphology based lung segmentation approach was presented. The other contribution is the presentation of the problem of nodule-type-identification and a naive approach for nodule type identification. In the absence of any groundtruth information on nodule types available, this work has presented the problem and solution approach in the light of the lung segmentation method. It has also discussed how analysis of lung segmentation in terms of nodule types may aid in measuring and examining the ability of a lung segmentation or nodule detection approach in terms of its

ability to identify specific types of nodules. With groundtruth, the nodule type identification approach can be further scrutinized and improved in future.

Acknowledgments

This work was carried out at ABV- IITM Gwalior, India with funding from DST, Govt. of India. The File Number is SR/WOSA/ET-153/2017

References

- [1] Adams, R. and Bischof, L. (1994) : Seeded region growing.. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 16, No. 6, pp. 641–647.
- [2] Armato III, S. G. and Sensakovic, W. F. (2004): Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis. *Academic Radiology*, Vol. 11, No. 9, pp. 1011–1021.
- [3] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., ...and Kazerooni, E. A. (2011) : The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, Vol. 38, No.2, pp. 915-931.
- [4] Cha, J., Farhangi, M. M., Dunlap, N. and Amini, A. A. (2018) : Segmentation and tracking of lung nodules via graph-cuts incorporating shape prior and motion from 4D CT. *Medical physics*, Vol. 45, No. 1, pp. 297–306.
- [5] Choi, W. J. and Choi, T. S. (2014) : Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Computer methods and programs in biomedicine*, Vol. 113, No. 1, pp. 37–54.
- [6] Dhara, A. K., Mukhopadhyay, S. and Khandelwal, N. (2012) : Computer-aided detection and analysis of pulmonary nodule from CT images: A survey. *IETE Technical Review*, Vol. 29, No. 4, pp. 265-275.
- [7] Farhangi, M. M., Frigui, H., Seow, A. and Amini, A. A. (2017) : 3-D active contour segmentation based on sparse linear combination of training shapes (SCoTS). *IEEE transactions on medical imaging*, Vol. 36, No. 11, pp. 2239–2249.
- [8] Gong, J., Liu, J. Y., Wang, L. J., Sun, X. W., Zheng, B. and Nie, S. D. (2018) : Automatic detection of pulmonary nodules in CT images by incorporating 3D tensor filtering with local image feature analysis. *Physica Medica*, Vol. 46, pp. 124–133.
- [9] Gupta, A., Saar, T., Martens, O. and Moullec, L. Y. (2018) : Automatic detection of multi-size pulmonary nodules in CT images: large scale validation of a multi-layer perceptron based false positive reduction step. *Med Phys*, Vol. 45, No. 3, pp. 1135–1149.
- [10] Hancock, M. C. and Magnan, J. F. (2016) : Lung nodule malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms: probing the Lung Image Database Consortium dataset with two statistical learning methods', *Journal of Medical Imaging*, Vol. 3, No. 4, pp. 044504.
- [11] Jamil, I. and Butt, S. I. (2016) : Adaptive thresholding technique for segmentation and juxtapleural nodules inclusion in lung segments. *International Journal of Bio-Science and Bio-Technology*, Vol. 8, No. 5, pp. 105–114.
- [12] Kumar, S. P. and Latte, M. V. (2019) : Fully automated segmentation of lung parenchyma using break and repair strategy. *Journal of Intelligent Systems*, Vol. 28, No. 2, pp. 275– 289.
- [13] Li, B., Christensen, G. E., Hoffman, E. A., McLennan, G. and Reinhardt, J. M. (2003) : Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. *Academic radiology*, Vol. 10, No. 3, pp. 255-265.
- [14] LObe and Lung Analysis 2011: <https://lola11.grand-challenge.org/Home/>, (Last Accessed September 2020).
- [15] Lung CT (2017): Data from lung CT segmentation challenge. The cancer imaging archive.
- [16] Mansoor, A., Bagci, U., Foster, B., Xu, Z., Papadakis, G. Z., Folio, L. R., ... and Mollura, D.J. (2015): Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends. *RadioGraphics*, Vol. 35, No. 4, pp. 1056–1076.
- [17] Mesanovic, N., Grgic, M., Huseinagic, H., Males, M., Skejic, E. and Smajlovic, M. (2011, February): Automatic CT image segmentation of the lungs with region growing algorithm. In *18th International Conference on Systems, Signals and Image Processing-IWSSIP*, pp. 395–400.
- [18] Mukhopadhyay, S. (2016) : A segmentation framework of pulmonary nodules in lung CT images. *Journal of digital imaging*, Vol. 29, No. 1, pp. 86-103.
- [19] Naqi, S. M., Sharif, M. and Yasmin, M. (2018) : Multistage segmentation model and SVM- ensemble for precise lung nodule detection. *International journal of computer assisted radiology and surgery*, Vol. 13, No. 7, pp. 1083–1095.
- [20] Netto, S. M. B., Silva, A. C., Nunes, R. A. and Gattass, M. (2012) : Automatic segmentation of lung nodules with growing neural gas and support vector machine. *Computers in biology and medicine*, Vol. 42, No. 11, pp. 1110–1121.
- [21] Sellathamby, Dinesh, (2017) : Lung Segmentation Data Science Bowl 2017, <https://www.kaggle.com/dineshsellathamby/lung-segmentation/>.
- [22] Setio, A. A., Jacobs, C., Gelderblom, J. and van Ginneken, B. (2015) : Automatic detection of large pulmonary solid nodules in thoracic CT images. *Medical physics*, Vol. 42, No. 10, pp. 5642-5653.
- [23] Sluimer, I., Schilham, A., Prokop, M. and Van Ginneken, B. (2006) : Computer analysis of computed tomography scans of the lung: a survey, *IEEE transactions on medical imaging*, Vol. 25, No. 4, pp. 385–405.
- [24] Wei, Y., Shen, G. and Li, J. J. (2013) : A fully automatic method for lung parenchyma segmentation and repairing. *Journal of digital imaging*, Vol. 26, No. 3, pp. 483-495.
- [25] Xiao, X., Zhao, J., Qiang, Y., Wang, H., Xiao, Y., Zhang, X. and Zhang, Y. (2018): An automated segmentation method for lung parenchyma image sequences based on fractal geometry and convex hull algorithm. *Applied Sciences*, Vol. 8, No. 5, pp. 832.
- [26] Zhang, S., Zhao, Y. and Bai, P. (2018): Object Localization improved GrabCut for Lung Parenchyma Segmentation. *Procedia computer science*, Vol. 131, pp. 1311–1317.
- [27] Zhang, W., Wang, X., Li, X. and Chen, J. (2018) : 3D skeletonization feature based computer- aided detection system for pulmonary nodules in CT datasets. *Computers in biology and medicine*, Vol. 92, pp. 64–72.

Authors Profile



R. Jenkin Suji received her Bachelors in Electronics and Instrumentation Engineering and Masters in Applied Electronics from Anna University, Chennai. She is currently a WOSA researcher at ABV-IIITM Gwalior. Her research interests are in the area of Biomedical image processing and computer Vision.



Sarita Singh Bhadauria is currently a senior faculty in the Electronics Department of RGPV University, Bhopal and is currently on lien from the Department of Electronics & Communication Engineering, Madhav Institute of Technology & Science Gwalior. Her research areas include image processing, adhoc-networks and software engineering