

# A Novel Hybrid Segmentation and Refinement Method for Automatic Lung Cancerous Nodules Extraction

P. Samundeeswari, R. Gunasundari

**Abstract**— The doctors uses chest Computed Tomography (CT) images to manually analyze the presence of cancerous nodule during cancer screening process. Due to heterogeneous and low intensity nature of CT image, manual image analyzing becomes difficult which leads to different problems like false positive detection, consumption huge analyzing time, observer error, etc. Developing an efficient automatic Computer Aided Detection (CAD) system is the most efficient approach to reduce the frequency of missed lung cancer and to make diagnosis simpler and time saving. The CAD system improves the accuracy of lung tumor detection and survival rate of the patient. In this paper, a fully automated model is presented for NSCLC nodule(s) segmentation from CT scan image. The proposed method follows three steps: (1) Preprocessing, (2) Automatic Lung Parenchyma Extraction and Border Repair (ALPE&BR) and (3) Automatic lung nodules segmentation using Connected Component Analysis (CCA) and Threshold Based Mathematical Nodule (TBMN) refinement algorithm. The ALPE&BR method consists of Automatic Single Seeded Region Growing (ASSRG) Algorithm for automatic lung parenchyma extraction and novel border concavity closing algorithm to get clear lung boundary. The proposed method successfully segments the true cancerous nodules by filtering out false region such as vessels, bone, fat, soft tissues, etc. The proposed method can provide the SN of 99.4%, SP of 98.5%, FPR of 0.6%, DSC of 0.982 and accuracy of 98.8%. These results are used to demonstrate that the proposed method outperforms the existing lung nodule segmentation method.

**Keywords**— Contrast limited adaptive histogram equalization, lung CT image, Automatic lung cancer nodule segmentation, wiener filter, grow cut algorithm, border concavity closing, automatic single seeded region growing algorithm.

## I. INTRODUCTION

Lung cancer seems to be one of the most common causes of death among people throughout the world. The American Cancer Society estimates the numbers of new cancer cases and deaths occurring in the United States every year by sex wise[5]. In 2017, they estimated that the lung and bronchus cancer occupied the second place with new cancer cases for both male (116,990 cases, 14%) and female (105,510 cases, 12%). The lung and bronchus cancer causes highest death rate for both male (84590 cases, 27%) and female (71,280 cases, 25%) in the United States.

**Revised Manuscript Received on April, 07, 2019**

**P.Samundeeswari**, Research scholar, Dept. of electronics and communication engineering, Pondicherry engineering college, Puducherry, India.

**Dr.R.Gunasundari**, Professor, Dept. of electronics and communication engineering, Pondicherry engineering college, Puducherry, India.

In India, 0.95 million new cancer cases and 0.63 million cancer deaths occur every year [4]. The doubled cancer cases emphasized that cancer burden in India could be projected to increase substantially from about 1 million new cases in 2012 to more than 1.7 million per year by 2035. The majority of cases (~60%) are diagnosed after the disease has spread to distant areas and the 5-year survival rate for those with advanced disease is only around 15-19%. However, if detected early, when the tumor is still localized, the 5-year survival rate is much higher at 70%-90% with the likelihood of many of these cases being cured [6]. Hence, early cancer detection can increase the possibility of survival rate. There are two main types of lung cancer: Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). Early diagnosis of NSCLC is very essential because 80% to 85% of lung cancer cases are NSCLC and found in non-smokers. Doctors utilize some imaging technique to diagnose lung tumor, for example Magnetic Resonance Imaging (MRI), X-rays, Computed Tomography (CT) scan, Positron Emission Tomography (PET) scan etc. Low Dose Computed Tomography (LDCT) scans of the chest provide more detailed pictures than chest x-rays and are better for finding small abnormal areas in the lungs [17]. An important medicolegal challenge for the radiologists is to discriminate lung lesions (nodules) from mediastinal structures, pulmonary vessels, bones and other complex anatomical regions on chest CT image. The reasons for a misdiagnosis on chest radiography and less frequently on CT scans, are many but often related to observer error (scanning, detection and decision-making errors), specific characteristics of the unobserved lesion (conspicuity, size and location) and technical imprecision [7]. The manual analysis becomes difficult and time-consuming and it leads to exhaustion besides reducing concentration duration which can adversely affect the radiologist diagnosis. Developing an efficient automatic Computer Aided Detection (CAD) system is the most efficient approach to reduce the frequency of missed lung cancer, to make diagnosis simpler and timesaving. The CAD system also improves the lung tumor detection accuracy and the survival rate of patient. The rest of the paper is organized as follows: The related work of the existing methods for segmentation of lung parenchyma and nodule regions is given in section 2. Section 3 deals with the proposed automated model for NSCLC nodule(s) segmentation from the CT scan image. The experimental results and discussion are given in Section 4. Finally, conclusion and future work are presented in section 5.



## II. RELATED WORK

A huge amount of investigation has been carried out on the topic of lung parenchyma and nodule segmentation in chest CT scans. In general, CAD systems consist of the following stages: 1) image preprocessing, 2) lung parenchyma and nodule segmentation and 3) False Positive (FP) rate reduction. The most common and simplest method for lung region segmentation is Thresholding [21, 20]. Due to heterogeneous intensity nature of chest CT scan image, it is difficult to identify the single accurate threshold intensity value of the lung region as distinguished from the surrounding structures.

Level set segmentation [8, 19] method is region based but faster, more robust and accurate. But it is not fully automatic; it needs prior knowledge about texture pattern. The values of the upper and the lower thresholds are selected manually and better selection of the threshold value provides superior output. Juanjuan Zhao *et al.* [13] aimed to isolate the lungs from the surrounding structures by improved region growing method. Here region growing algorithm was applied to segment the lung and then improved eight neighbor region growing was employed to remove noise. They utilized corrosion and expansion operations for lung boundary smoothening. In [12, 14], the segmentation algorithm was applied based on thresholding and watershed segmentation to detect the cancerous cell or nodule of a lung CT scan image.

Mehdi Alilou *et al.* [9] developed the application of a gradient vector flow active contour model for nodule boundary extraction with manual seed selection. Ezhil *et al.* [11] developed a region based active contour model to separate the parenchyma and reconstructed it using Selective Binary and Gaussian Filtering with new Signed Pressure Force function (SBGF-new SPF). Then Fuzzy C-Means (FCM) technique was used to segment lung nodules. Wei Shen *et al.* [16] created an end-to-end architecture of deep learning, named Multi-Crop Convolutional Neural Networks (MC-CNN) used to extract nodule features directly and classify the suspicion malignancy from CT scan image with 87.14% accuracy rate.

Xiang-Xia Li *et al.* [3] proposed an improved random walker method to automatically segment nodules from lung CT scan image. They utilized the manually cropped pulmonary nodule region marked by specialists to select the seed point automatically. The geodesic distance was applied to find nodule center where as nodule foreground seeds were sampled from the circle with radius (R) and background seeds were sampled with the circle of radius 4R. The improved RW algorithm was able to segment various types of pulmonary nodules and attained a desirable segmentation quality.

Heewon Chung *et al.* [2] adapted the Chan–Vese (CV) model for active contour method followed by a Bayesian approach to detect and segment juxta-pleural nodule candidates. They analyzed consecutive slices of the same lung CT images to find proper lung contour region. They employed concave points detection and circle/ellipse Hough transforms to eliminate false positives. Hongyang Jiang *et al.* [1] implemented automatic lung nodule detection based on multi-group patches slice out from the lung CT images. Multi-scale Frangi filter was used to eliminate vascular

structures and give nodules. Through combining nodule cropped from original image and cropped binary images after the vessel-eliminated lung, a four-channel Convolution Neural Networks (CNN) model was designed to detect four levels of nodules by learn the knowledge from radiologists based on diameter.

In this research, an automated model is proposed for NSCLC nodule(s) segmentation from CT scan image. The proposed method is divided into three stages: (1) Preprocessing, (2) Automatic Lung Parenchyma Extraction and Border Repair (ALPE&BR) and (3) Automatic lung nodules segmentation using Connected Component Analysis (CCA) and Threshold Based Mathematical Nodule (TBMN) refinement algorithm to filter out false nodule region such as vessels, bone, fat, soft tissues, etc., and extract segmented true cancerous nodule, as shown in Fig.1.

## III. MATERIALS AND METHODS

### A. Chest CT images database Collection:

In the present study, LIDC/IDRI is a public database [22] that contains a total of 1018 helical thoracic CT scans obtained from 1010 different patients. It consists of 512\*512 image size with pixel dimension range of 0.51 to 0.74mm. The process of annotating the nodules of the LIDC/IDRI database was performed by four radiologists in two stages. The databases were collected from different kinds of imaging devices, with the standardized mean pixel dimension is 0.688mm considered for the proposed work. The nodules' sizes may be either too small or too big and had different effect on detection[22]. In this study, 30 thoracic CT scan images were selected with different nodule types (solid, partially solid or non-solid), dissimilar texture (shape, size, area and perimeter) and different locations (isolated, multinodule, juxtavascular and juxtapleural) to evaluate the effectiveness of the proposed method.

### B. Preprocessing:

Image preprocessing is one of the most essential steps to ensure the high precision of the nodule detection process. In medical image processing, it is very important to preprocess the image so that automatic segmentation and feature extraction algorithms work properly. This paper developed a new pre-processing approach for lung CT image which is shown in Algorithm 1. This approach enables precise lung parenchyma segmentation process by eliminating artifacts. The exact lung parenchyma region segmentation ensures improvement in the accuracy of lung nodule segmentation. In preprocessing, initial step is to map the gray levels of grayscale CT input image (Fig.2(a)) with hot color image in all the three layers (Red, Green, and Blue) shown in Fig.2(b). The second step is to separate individual (R, B & G) color components and this improves the edge separation between nodules attached with mediastinum.

The red color component is not considered in this work since it has very high intensity and does not preserve edge details in heterogeneous nature of the CT image. The edge

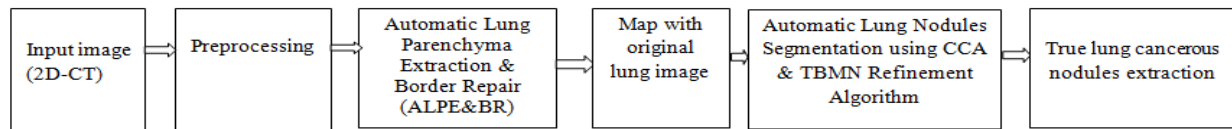


Fig.1 Block diagram of proposed method

detail is very essential for segmentation process to separate lung parenchyma from the surrounding regions such as mediastinum, fat, heart, bone, thoracic wall, etc. Hence, we are interested to proceed the further process with either B or G component in their grayscale intensity.

Algorithm 1: Preprocessing

Input: Lung 2D-CT gray scale image (I)

Output: Background removed image

1. Convert I into color image
2. Separate (G & B) color components
3. Convert individual component into gray scale (Gg & Bg).
4. Enhance Bg component using CLAHE
5. Calculate total no. of pixel ( $\Psi$ )
6. Calculate sum of the pixel intensity ( $\Theta$ )
7. If  $\Theta > (1/4) * \Psi$   
 $I_g = \text{Enhanced (Bg)}$   
 Else  
 $I_g = \text{Enhanced (Gg) using CLAHE}$
8. Perform wiener filtering for  $I_g$
9. Remove the background using Grow cut algorithm.

The selection of either Bg or Gg component is explained in algorithm 1. Our main interest is to proceed the process with Bg (Fig.2(d)) because it preserves edge separation detail between lung and mediastinum region better than Gg (Fig.2(c)). These details will make simple and efficient CAD system analysis for cancerous nodule invading the mediastinum region. But for very low contract CT image, Bg component does not contain even minimum necessary structure details. Hence, in that case Gg is selected instead of Bg and is followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) [15] technique is adapted to enhance the contrast of either Bg or Gg based on step 7 output (Alg.1). CLAHE operates on tiles (small regions) instead of the entire image. It enhances each tile's contrast and combines neighboring tiles using bilinear interpolation to reduce artificially induced boundaries and is shown in Fig.3(a).

Wiener filter [18] is basically the optimum linear filter which adapts itself to the local image variance, inverts the blurriness, and reconstructs the degraded image. It is employed to remove additive noise while preserving the fine structure details and edges. The filter with size  $5 \times 5$  is selected, it creates pixel-wise linear filtering by estimating local mean ( $\mu$ ) and variance ( $\sigma$ ) from a local neighborhood of each pixel.

$$F(m, n) = \mu + \frac{(\sigma^2 - v^2)}{\sigma^2} (I(m, n) - \mu) \quad (1)$$

where  $I$  and  $F$  denote the original and Wiener filtered image and  $V$  denotes noise variance. The wiener filter output is shown in Fig.3(b) given as an input to background removal process using GrowCut segmentation and morphological holes filling algorithm.

and its state is stored as a three-tuple (l, s, C), with l as the foreground/background label, s as the strength of the cell and C encoding the gray scale information of the corresponding pixel.

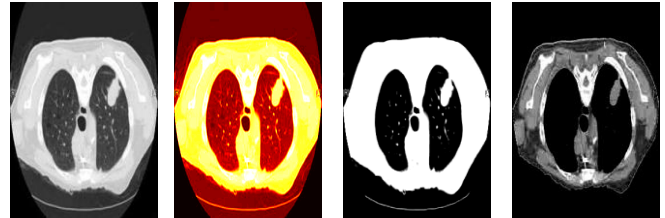


Fig.2 a) Input image (2D-CT) b) Color mapped output c) G component d) B component

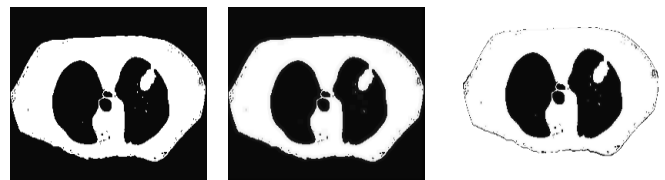


Fig.3 a) CLAHE output b) Wiener filter c) Background removed image

In this work is proposed an automatic foreground and background labels selection, by using the advantage of lung CT image structure. The lung CT image structure usually contains high intensity in its middle because of the presence of mediastinum between the right and the left lung lobes. Automatic foreground labels are selected by finding the central pixel and form clustering of its four neighbor pixels, it is considered as the foreground with labeling  $l=1$ . Automatic background labels are selected by choosing four corner pixels in the image, because it is usually dark region and label  $l=0$ . After the initialization, this algorithm iteratively labels all the remaining pixels either as background or foreground and terminates when all pixels have been assigned a background or foreground label. Furthermore, a morphological hole filling is applied to fill the holes due to lung lobe regions and after that is obtained a mask image with foreground and background labels. Finally, background removed image is obtained by performing mathematical XOR operation between mask and Wiener filter output.

C. Automatic Lung Parenchyma Extraction & Border Repair (ALPE&BR)

Lung segmentation is an important prerequisite step for automatic chest CT image analysis. Lung segmentation involves isolating the lungs by removing the tissue outside parenchyma and identifying the lung boundary perfectly.



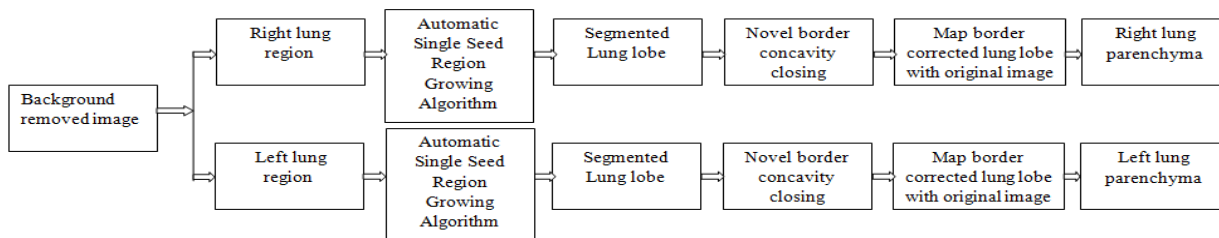


Fig.4 Proposed ALPE&BR Block diagram

The nodules appear within lung parenchyma at different locations like main bronchus, mediastinum, lobar bronchus, pleura etc. In automatic lung nodule detection process, if lung segmentation fails to correctly extract the borders of the lung’s wall, the nodules attached with the borders may be missed out partially or fully. Thus, automatic and accurate lung segmentation algorithm is required. To perfectly segment the lung parenchyma, an automatic scheme is proposed in this research. The procedure for the proposed ALPE&BR can be seen in Fig.4. It considers five main steps. In the first step, acquire preprocessing output, then separate right and left lung region by equally dividing the image which is shown in Fig.5 a(i)&b(i). This individual region analyzing is required to avoid over/under segmentation effects and make automatic seed selection processes easier. The automatic lung lobe segmentation and border refinement steps are described in the next subsections.

1) Automatic Single Seeded Region Growing(ASSRG) Algorithm:

In general lung lobe region is filled with air, hence, this region is low contrast and homogenous compared with the surrounding region. In this research, Region growing segmentation method is employed to segment right/ left lung lobe region with the proposed automatic single seed selection procedure. With the common knowledge about lung intensity, shape, and position, an automatic single seed selection is proposed and lung lobe region extraction is explained in Algorithm 2. This algorithm segments lung lobe with correct boundary for isolated type nodules. But for nodule type juxtavascular or juxtapleural, it generates lobe boundary with concavity in the nodule position and is shown in Figs.5 a(ii)&b(ii).

Algorithm 2: ASSRG

- Input: Image (I) either right or left lung region  
 Output: Segmented lung lobe (right or left) region
- 1: Automatically select single seed pixel position and intensity in image (I)
    - a. Input image (I) either right or left lung region
    - b. Find the center pixel
    - c. Check if that dark region (intensity threshold<40) denotes air region in lung
    - d. Again check if its maximum number of neighbors are dark region
    - e. If it is dark region, then consider that pixel as seed pixel position and intensity
    - f. Else iteratively grow the center pixel towards all four directions
    - g. Repeat step (c), (d) and (e)
  - 2: Check the neighboring pixels using  $|neighbor\ pixels - seed| < Threshold$  and add them to the region if they are similar to the seed
  - 3: Grow the seed regions by including adjacent pixels that satisfy the similarity rule
  - 4: Repeat step 2 for each of the newly added pixels; stop if no more adjacent pixels to be included in a seed region.
  - 5: Output is segmented lung lobe (right or left) region

or juxtapleural nodules. Ignoring boundary concavity leads to missing out juxtavascular or juxtapleural nodules extraction.

Algorithm 3: Novel border concavity closing algorithm

Input: Region growing segmented Lung lobe (J), Line width=n  
 Output: Border corrected lung lobe (I)

1. Insert ‘n’ columns of zeros to left and right of J. % (Improves boundary detection effectiveness)
2. Morphologically opening operation removes unwanted objects on image
3. Morphological reconstruction using flood-fill operation to fill holes in images
4. Regularizing image shape by applying median filter
5. Apply connected components analysis
6. Find no. of connected region (N)
7. for j=1:N  
 Calculate Area (j) of each connected region  
 end
8. Extract connected region (I) which has maximum Area. % Lung lobe region
9. Find all indexes of I which has true intensity and store it as  $\phi=(x(i),y(i))$ . % where i= no. of positions having true intensity
10. Apply boundary detection algorithm [23] with indexes set  $\phi$ , which generates a set of boundary indexes ( $\chi$ ). % Except concave indexes.
11. Connect indexes ( $\chi$ ) with line width (n) which completely enclosing lung lobe without concavities
12. Apply canny edge to output of step 10 and find edges
13. Repeat step5 to step 7
14. Extract connected region which has minimum Area (Imin). % (interior lung lobe edge without presence of over segmented outer soft tissue wall region)
15. Fill the interior region of Imin with high intensity and remove step1 inserted extra columns
16. Output is border corrected lung lobe (I)

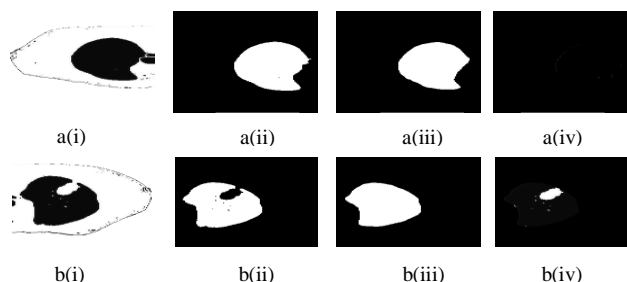


Fig.5 ALPE&BR results: a(i)&b(i) Right and left lung after background removal, a(ii)&b(ii) ASSRG produced lung parenchyma, a(iii)&b(iii) After border concavity closing, a(iv)&b(iv) Extracted right/ left lung lobe is superimposed with original right/ left CT image.

Thus, a novel border concavity closing algorithm is proposed, which is explained in Algorithm 3. The proposed algorithm make uses of morphological operation, connected component analysis, logical rules to extract lung lobe region based on its area;

then is applied the boundary detection



algorithm to collect boundary indexes by excluding deep concave indexes and employed clockwise line connection between collected boundary indexes with some line width ( $n \geq 4$ ) which produces thick line boundary without concavity.

Next, canny edge detection is employed to detect inner and outer edge of the thick boundary line. To avoid over segmented lung wall region, inner edge of the thick line is extracted by calculating area (minimum) which encloses the entire lung lobe regions. This proposed algorithm successfully provides lung lobe region with proper border without concavity and over segmented lung walls/mediastinum is shown in Figs.5 a(iii)&b(iii). In this section, finally the extracted right/ left lung lobe is superimposed with the original right/ left CT image as shown in Figs.5 a(iv)&b(iv). This proposed technique ensures the absolute isolation of juxtavascular or juxtapleural nodules from the outer wall/ mediastinum region.

**D. Automatic Lung Nodules Segmentation by CCA & TBMN Refinement Algorithm**

The right/ left lung parenchyma may contain the presence of cancer nodules which may also include some other findings such as pulmonary nodule, Ground Glass Opacity (GGO), vessels, bronchioles and other impurities. The presence of these other findings in the parenchyma affects the performance in terms of lung cancerous nodules detection accuracy. The aim of this section is to eliminate the small non-cancerous nodules (<1cm), vessels structures and other impurities in the lung parenchyma, thus we can beneficially enhance analyze of cancer nodule. Thus in this study an automatic lung nodules segmentation algorithm is proposed using Connected Component Analysis (CCA) and Threshold Based Mathematical Nodule (TBMN) refinement; it is explained in Fig.6.

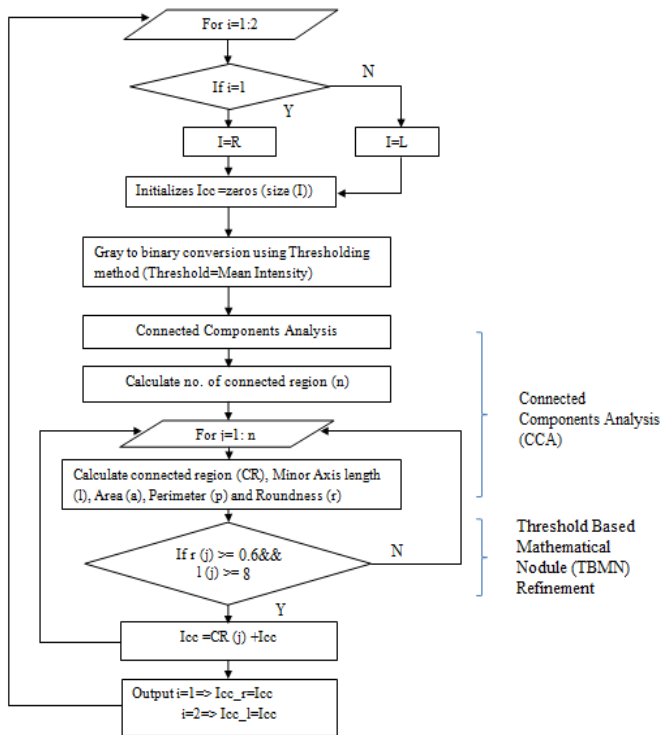


Fig.6 Flow diagram of proposed automatic lung nodules segmentation algorithm

the image is converted into binary; then CCA is applied to find the number of Connected Regions (CR) in I, for every

CR minor axis length (l), area (a), perimeter (p) and roundness (r) are found. To eliminate other surplus findings present in parenchyma, TBMN refinement analysis is applied based on minor axis length (l) and roundness. In TBMN refinement, first is analyzed roundness (r) property of each CR, because usually most of the cancerous nodules are closely related to circular shape. For circle shape the values of  $r=1$  and  $r=0$  denote elongated shape; 'r' value is utilized to eliminate vessels, bronchioles and other impurities as shown in Fig.6. Next process is to eliminate small non-cancerous nodule by analyzing dimension (minor axis length (l)) for every CR which has  $r > 0.5$ , value of 'l' that should satisfy minimum 1cm for cancerous nodule. The CR is extracted if its 'l' is greater than 1cm which means that it eliminates small non-cancerous nodule. Figs.7 (a&b) show that the proposed segmentation algorithm successfully provides nodules (i.e., CR's whose  $r > 0.5$  and  $l \geq 1$ cm) present in the right and the left parenchyma. The final output is the automatically segmented true cancerous nodule present in the right lung parenchyma (Icc\_r) and the left lung parenchyma (Icc\_l) are concatenated and mapped with original CT lung image is shown in Fig.7 (c).

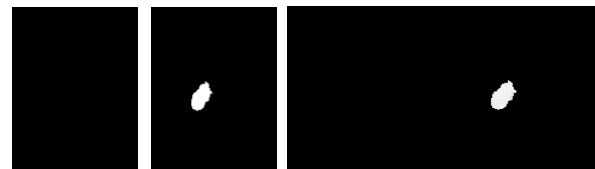


Fig.7 a) Cancerous nodules in right lung, b) Cancerous nodules in left lung, c) output of proposed automatic lung nodules segmentation in both lungs.

**IV. EXPERIMENTAL RESULTS AND DISCUSSION**

All our implementations and investigations were programmed in MatLab 2018a and executed on a personal computer configured with a 64 bit operating system, 2.53-GHz Intel Core i3 M 380 processor with 4 GB of RAM. In this study, 30 images were selected from LIDC/IDRI public database, each having 512\*512 image size. For experimental analysis image dimension was resized into 256\*256, with each pixel length range being 1.02 to 1.48mm. The database was collected from different kinds of imaging devices; the standardized mean pixel dimension 1.376mm was considered for the proposed work. The experimental step by step outcomes of the proposed method are discussed in this section. Figs.8 & 9 show the proposed preprocessing algorithm outputs. Fig.8 (a) shows the lung CT image with solitary cancerous nodule. The RGB color mapping and separation of G & B gray scale components are shown in Figs.8 (b, c and d). Here, B component is selected for further process because of its total intensity  $> (1/4) * \text{image size}$ . Figs.9 (a and b) gives the output of CLAHE and Wiener filter which preserve the image details. After this, grow cut algorithm was applied to remove the background and extract the foreground lung region shown in Fig.9 (c).



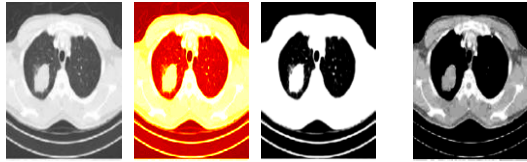


Fig.8 a) Input image (2D-CT) b) Color mapped output c) G component d) B component



Fig.9 a) CLAHE output b) Wiener filter c) Background removed image

Fig.10, explains the step by step process involved in Automatic Lung Parenchyma Extraction & Border Repair (ALPE&BR) process. To avoid over/under segmentation effects background removed outcome was equally divided into right and left lung as shown in Figs. 10 (a(i)&b(i)). Right/left lung parenchyma region was segmented by applying the proposed ASSRG algorithm as shown in Fig.10 a(ii)&b(ii) which automatically extracted lung region from its surrounding and it was followed by novel border concavity closing algorithm being applied to close concave regions in the outer lung lobe; this step was mainly imperative to find nodules attached with the lung wall region being shown in Figs.10 a(iii)&b(iii). Finally, lung parenchyma was successfully segmented by the mapping of original lung (right/left) image with concavity closed border image.

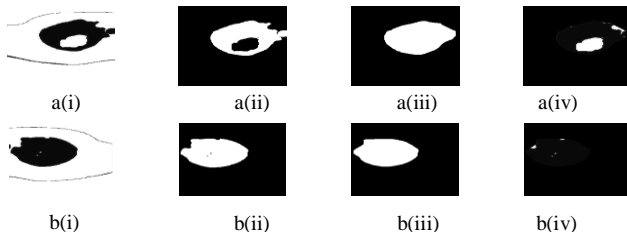


Fig.10 ALPE&BR results. a(i)&b(i) Right and left lung after background removal, a(ii)&b(ii) ASSRG produced lung parenchyma, a(iii)&b(iii) After border concavity closing, a(iv)&b(iv) Extracted right/ left lung lobe is superimposed with original right/ left CT image.

In the proposed method, the main objective is to identify and segment the true cancerous nodules in given CT lung image and to achieve this aim automatic lung nodules segmentation by CCA & TBMN refinement Algorithm is proposed. Here CCA is applied to find each Connected Region (CR) and its parameters such as area(a), perimeter(p), minor axis length(l) and roundness(r) are calculated. For individual CR, area denotes total number of pixel, perimeter denotes distance around the boundary, and minor axis length denotes the length of the minor axis. Then roundness nature of each CR is calculated by using eq.(2).

$$r = \frac{4 \cdot \pi \cdot a}{p^2} \quad (2)$$

To segment and extract the true cancerous nodules, Threshold Based Mathematical Nodule (TBMN) refinement technique is applied. It checks if the value of 'l' is (should be) greater than 8 pixel (i.e. 1cm) and r is (should be) greater than 0.5 for each CR. Those CRs that satisfy the above criteria are considered as true cancerous nodules or else it may be some other impurities (e.g. vessels, bronchi, etc.). By this way the proposed method efficiently eliminates false nodules and extracts true cancerous nodules. Figs.11 (a)&(b) shows true cancerous nodules in

the right and the left lung region, the proposed automatic true cancerous nodule segmentation outcome is shown in Fig.11 (c). Fig.12 shows the proposed method's performance well with all nodule types, shape, and texture.

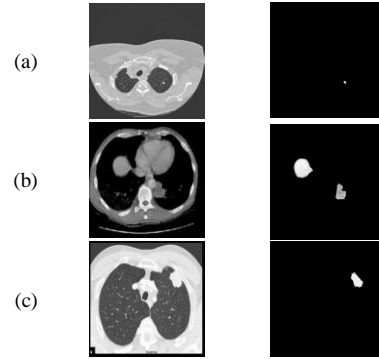


Fig.12 Sample output for different nodule types. a) Solitary nodule b) nodule invades mediastinum region c) Juxtapleural

#### A. Quantitative analysis

To evaluate the proposed method's performance, several statistical parameters were used. They were Sensitivity (SN), Specificity (SP), False Positive Rate (FPR), Dice Similarity Coefficient (DSC) and Accuracy. To compute these metrics, first the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values were calculated by comparing radiologist manually segmented region ( $S_M$ ) and the proposed method's automatically segmented region ( $S_A$ ). TP and FP are the numbers of positive pixels (nodule region) labeled correctly and incorrectly. TN and FN are the numbers of negative pixels (non-nodule region) labeled correctly and incorrectly.

$$TP = \frac{S_A \cap S_M}{S_M} \quad (3)$$

$$TN = \frac{S_A \cup S_M}{S_M} \quad (4)$$

$$FP = \frac{S_A - S_M}{S_M} \quad (5)$$

$$FN = \frac{S_M - S_A}{S_M} \quad (6)$$

We briefly explain these metrics, SN and SP measure the ability of the proposed algorithm to correctly segment the cancerous nodule and non-cancerous nodule, FPR measures the ability of the false positive detection rate of the proposed algorithm, DSC is a statistical approach used to compare the similarity of  $S_A$  and  $S_M$ . The FPR value ranges between 0 and 100 in percentage, where 0 means there is no false positive region segmented and 100 means there is highly false positive segmented region. The DSC value ranges between 0 and 1, where 0 means there is no similarity and 1 means there is perfect similarity. The accuracy measures the reliability of the proposed method to detect cancerous nodules from lungs region. The performance metrics such as SN, SP and accuracy range from 0 to 100, with 0 signifies poor performance and 100 signifies high performance.



These ranges of values were used to evaluate the effectiveness of the proposed method.

$$SN = \frac{TP}{TP+FN} * 100 \quad (7)$$

$$SP = \frac{TN}{TN+FP} * 100 \quad (8)$$

$$FPR = 100 - SN \quad (9)$$

$$DSC = \frac{2*TP}{2*TP+FN+FP} \quad (10)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (11)$$

Finally, the performance of the proposed method was evaluated by comparing with the existing method is shown in Table 1. Automatic cancerous nodule segmentation with the proposed method on all lung CT images ( $N = 30$ ) was performed. The proposed method has provided the highest sensitivity, specificity, disc similarity coefficient and accuracy values are 99.4, 98.5, 0.982 and 98.8, than the other existing methods. Regarding FPR, the mean value (0.6) was much lower what was in the other methods. The proposed method performs well with all types of nodule (smooth and regular mediastinum, solitary, juxtavasular and juxtapleural) than other existing method, but still it has a minor limitation in mediastinum region nodules segmentation, which is due to its complex irregular structure. While applying concavity closing algorithm small piece of mediastinum regions were segmented and it looks similar to the cancerous nodules by satisfying the criteria's ( $r > 0.5$  and  $l \geq 1cm$ ) which leads to false positive nodule segmentation if nodule invades complex mediastium. In future, we plan to overcome the above drawback by designing a novel nodule filter using Hounsfield unit value and this will improve the accuracy and DSC of the segmentation method.

TABLE 1: Performance comparison

Researches	SN (%)	SP (%)	FPR (%)	DSC	Acc (%)
Bayesian approach based on the Chan-Vese (CV) model [2]	97.1	96.3	-	0.97	96.6
OTSU and Convolution Neural Network for nodule detection [1]	(80-94)	-	(4.7-15.1)	-	-
Proposed Method	99.4	98.5	0.6	0.982	98.8

## V. CONCLUSION

A novel lung cancerous nodule segmentation and extraction algorithm capable of detecting solitary, multi-nodule, juxta-vascular and juxta-pleural nodules was proposed. The proposed method was based on steps such as preprocessing which enhanced nodule detection process by eliminating artifacts and boundary. Then lung parenchyma extraction was performed using ALPE&BR algorithm which improved the accuracy of juxta-vascular and juxtapleural nodules segmentation. Finally, automatic lung nodules segmentation using Connected Component Analysis (CCA) and Threshold Based Mathematical Nodule (TBMN) refinement algorithm which successfully segmented the true cancerous nodules by filtered out other impurities. The proposed method outperformed the other existing methods in terms of providing highest value with SN of 99.4%, SP of 98.5%, FPR of 0.6%, DSC of 0.982 and accuracy of 98.8%. Therefore the proposed method will aid

radiologists in their interpretation of lung CT images, particularly for cancerous lung related quantitative analysis and it will help the doctors in early treatment planning which will increase the survival rate of patient.

## FUTURE WORK:

In future, we aim to eliminate false nodules by designing a nodules filter using Hounsfield Unit (HU) value. Then we plan to develop an automatic classifier to classify the lung cancer sub-types and stages by extracting features from the segmented nodules.

## REFERENCES

- Hongyang Jiang, He Ma, Wei Qian, Mengdi Gao and Yan Li, "An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network", *IEEE Journal of Biomedical and Health Informatics*, Vol.22, Issue.4, pp.1227-1237, July 2018.
- Heewon Chung, Hoon Ko, Se Jeong Jeon, Kwon-Ha Yoon and Jinseok Lee, "Automatic Lung Segmentation with Juxta-Pleural Nodule Identification using Active Contour Model and Bayesian Approach", *IEEE Journal of Translational Engineering in Health and Medicine*, Vol.6, ISSN: 2168-2372, May 2018.
- Xiang-Xia Li, Bin Li, Lian-Fang Tian and Li Zhang, "Automatic benign and malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm", *IET Image Processing*, Vol. 12, Issue. 7, pp. 1253-1264, Jan. 2018.
- Krishnaveni K, Rosmi Jose, Sumitha SK, Teena Johny, Shanmuga Sundaram R and Sambathkumar R, "A Study on Socio Demographic and Associated Risk Factors for Cancer Patients in Private Cancer Hospital, Bangalore, India", *Research Journal of Pharmacy and Technology*, Vol.11, Issue.2, pp.677-680, 12 June 2018.
- Rebecca L. Siegel, Kimberly D. Miller and Ahmedin Jemal, "Cancer Statistics, 2017", *CA: A Cancer Journal for Clinicians*, Vol. 67, No.1, pp.7-30, Jan 2017.
- Sean Blandin Knight, A. Crosbie, Haval Balata, Jakub Chudziak, Tracy Hussell and Caroline Dive, "Progress and prospects of early detection in lung cancer", *The Royal Society Publishing, Open Biology*, Vol.7, 6 sep 2017.
- Annemilia del Ciello, Paola Franchi, Andrea Contegiacomo, Giuseppe Cicchetti, Lorenzo Bonomo and Anna Rita Larici, "Missed lung cancer: when, where, and why?", *Diagnostic and Interventional Radiology*, Vol.23(2), pp.118-126, Mar 2017.
- Mustafa Mohammed Sattar and N.Anupama, "Automated Detection and Classification of Cancerous Lung Nodule using Level Set Segmentation and ANN", *International Journal of Engineering Science and Computing*, Vol.7, Issue.5, pp.11341-11346, May 2017.
- Mehdi Alilou, Niha Beig, Mahdi Orooji, Prabhakar Rajiah, Vamsidhar Velcheti, Sagar Rakshit, Niyoti Reddy, Michael Yang, Frank Jacono, Robert C. Gilkeson, Philip Linden and Philip Linden, "An integrated segmentation and shape-based classification scheme for distinguishing adenocarcinomas from granulomas on lung CT", *Medical Physics*, Vol.44(7), pp: 3556-3569 July 2017.
- Jan Egger, Christopher Nimsky and Xiaojun Chen, "Vertebral body segmentation with GrowCut: Initial experience, workflow and practical application", *SAGE Open Medicine*, Vol. 5, pp. 1-10, 2017.
- Ezhi E. Nithila and S.S. Kuma, "Segmentation of lung nodule in CT data using active contour model and Fuzzy C-mean clustering", *Elsevier Alexandria Engineering Journal*, Vol. 55, Issue 3, Pages 2583-2588, Sep. 2016.
- S. Logesh kumar, M. Swathy, S. Sathish, J. Sivaraman and M. Rajasekar, "Identification of Lung Cancer Cell using Watershed Segmentation on CT Images", *Indian Journal of Science and Technology*, Vol.9(1), ISSN (Print) : 0974-6846, Jan. 2016.
- Juanjuan Zhao, Guohua Ji, Xiaohong Han, Yan Qiang and Xiaolei Liao, "An automated pulmonary parenchyma segmentation method based on an improved region growing algorithm in PET-CT imaging", *Springer Frontiers of Computer Science*, Vol.10, Issue 1, pp:189-200, Feb. 2016.
- Anam Quadri, Rashida Shujae and Nishat Khan, "Review on Lung Cancer Detection using Image Processing Technique", *International Journal Of Engineering Sciences & Research Technology*, Vol.5(2), ISSN: 2277-9655, pp.819, Feb. 2016.



15. Musthofa Sunaryo and Mochammad Hariadi, "Preprocessing on Digital Image using Histogram Equalization: An Experiment Study on MRI Brain Image", *International Journal of Computer Science and Information Technologies*, Vol. 7 (4), pp.1723-1727, 2016.
16. Wei Shen, Mu Zhou, Feng Yang, Dongdong Yu, Di Dong, Caiyun Yang, Yali Zang, Jie Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification", *Pattern Recognition*, vol. 61, 2016.
17. Geoffrey D Rubin, "Lung Nodule and Cancer Detection in CT Screening", *PubMed Journal of thoracic imaging*, Vol.30(2), pp. 130-138, Mar. 2015.
18. Shuangfeng Dai, KeLuan, JiyangDong, YifeiZhang and YongChen, "A novel approach of lung segmentation on chest CT images using graph cuts", *Neurocomputing, ScienceDirect*, Vol.168, pp. 799-807, 2015.
19. C. Li, R. Huang, Z. Ding, J. C. Gatenby, D. N. Metaxas, and J. C. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Transaction Image Processing*, Vol. 20, No. 7, pp. 2007-2016, 2011.
20. S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves and E. A. Kazerooni, "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, 2011.
21. J. Pu, J. K. Leader, B. Zheng, F. Knollmann, C. Fuhrman, F. C. Scirba and D. Gur, "A computational geometry approach to automated pulmonary fissure segmentation in CT examinations," *IEEE Transaction Medical Imaging*, Vol. 28, no. 5, pp. 710-719, May 2009.
22. S. Ukil and J. M. Reinhardt, "Anatomy-guided lung lobe segmentation in X-ray CT images", *IEEE Transaction Medical Imaging*, Vol. 28, no. 2, pp. 202-214, Feb. 2009.
23. <http://in.mathworks.com/help/matlab/ref/boundary.html>