Lung Segmentation Based on Statistical Analysis Using Features of Connected Components

V. Juliet Rani[1] · K. K. Thanammal[2]

Abstract

Lung diagnosis is one of the vital needs of the medical world. Lung cancer is the deadliest disease in the world which can be diagnosed using Computed Tomography (CT) images. The accuracy of lung cancer diagnosis in CT images through Computer Aided Diagnosis (CAD) system, depends on the accurate performance of lung segmentation method. Lung segmentation process extracts the lung region from the CT images, and it is challenged by the issues like less accuracy in segmentation, high false segmentation and high time consumption. Consequently, there is an essential necessity for a new lung segmentation method to resolve these issues, and to increase the performance of lung oriented CAD system. This paper proposes a novel lung segmentation method namely 'Lung Segmentation based on Statistical analysis using Features of Connected Components'. It performs the advanced statistical data processing on features of foreground area and connected components. A new approach of peak based analysis effectively extracts the true lung regions from the lung CT images. This method is enriched with morphological operations to gain high accuracy lung segmentation. It also delivers a new approach in left and right lung separation process via the Local Binary Pattern based texture processing which is a light weight algorithm to reduce time complexity. The performance analysis proves that this lung segmentation method is robust against scaling issues in lung CT images, and it absorbs less time consumption and noteworthy enhancement in segmentation-accuracy by achieving the value of 96.38% for lung segmentation.

Keywords CT lung segmentation · Statistical based lung segmentation · Medical image processing · Left and right lung separation · LBP

✉ V. Juliet Rani
  vjulietrani@gmail.com

  K. K. Thanammal
  thanaravindran@gmail.com

[1] Department of Computer Science and Centre for Research, S.T. Hindu College (Affiliated to Manonmaniam Sundaranar University, Tirunelveli), Nagercoil, Tamil Nadu, India

[2] Department of Computer Science and Centre for Research, S.T. Hindu College (Affiliated to Manonmaniam Sundaranar University, Tirunelveli), Nagercoil, Tamil Nadu, India

1 Introduction

Lung cancer is the deadliest disease prevailing all over the world. It can be cured if only diagnosed in the early stage. It can be easily detected using Computed Tomography (CT) images. Computer Aided Diagnosis (CAD) plays a key role in lung cancer treatment [10].

Lung segmentation is the computer-based process that can identify the boundaries of lung from surrounding thoracic tissue on CT images. It is a vital first step in diagnosing the lung diseases. Automatic lung segmentation methods using CAD systems are widely developed and this makes the clinical diagnosis efficient [5]. The lung segmentation process in CT images can transform the original image to more abstract and more closely packed than the earlier methods [9].

An opinion, based on medical experience, suggests that a significant improvement in lung imaging and computer technology leads to advances in radiation therapy planning and delivery, which includes techniques such as image-guided radiotherapy [27]. Lung segmentation is the critical initial step for the computer-aided detection and diagnosis system for lungs. The output of lung segmentation guides the lung cancer diagnosis system to act as an accurate CAD system, which helps pulmonologists for the early detection of cancer and preplanning to radiation treatment or surgery. Lung segmentation also reveals the damaged status of lungs due to pulmonary diseases, to pulmonologists. Left and Right lung separation provides an additional info about the damage of specific lung. In major existing methods, lung segmentation can be approached well only if the lungs exhibit minimal or no pathologic conditions [18]. Also, many lung segmentation methods have not support for left and right lung separation. The occurrence of false segmentation in lungs regions is a challenge in the lung segmentation process. The inadequate segmentation accuracy is another challenge in lung segmentation. Hence, a new lung segmentation method should coded to solve the challenges in lung segmentation process.

A. Morales Pinzon et al. describe a lung segmentation method based on the principles of cascade registration. The disadvantage of this method are high processing time and less accuracy [24]. Ratish chandra Huidrom et al. propose a fast thresholding-based lung segmentation method that extracts the lung mask from the CT scan image in a quick time span whereas the disadvantage of this method is low accuracy[11]. Geng Chen et al. describe an automatic pathological lung segmentation method which is used in low-dose CT images. The disadvantage of this method is the local shape reconstruction error [3]. Tao Peng et al. describe a hybrid lung segmentation on chest CT scans in which the lung boundaries can be easily detected [23].

The author Ahmed-Soliman et al. describe a lung segmentation method on CT chest images. This method uses adaptive appearance-guided shape modelling. But it is not effective if the healthy part of the lung is small [25]. Grigorios-Aris Cheimariotis et al. describe an automatic lung segmentation method using SPECT images. The disadvantage of this method is in locating the land marks in the left lungs [2]. Oluwakorede M. Oluyide et al. describe a lung segmentation method for CT chest images but it can not be applied in CAD system for lung cancer detection [21].

P.P.R Filho et al. describe a technique for segmentation of CT lung images. The disadvantage is that the edges of the lungs are not clearly defined [7]. The author Weihang Zhanga et al. describe an automated lung segmentation on CT images. This method is not good for the lungs which have diffuse parenchyma lung diseases and for lungs the lungs with long-term breathing problems [26]. Heewon Chung et al. describe an automatic lung segmentation method with juxta-pleural nodule identification. This method uses active

contour model and bayesian approach. The disadvantage of this method is that the nodule inside the lung parenchyma cannot be identified [4].

Rizki Nurfauzi et al. describe a lung segmentation method using adaptive border method. The disadvantage of this method is that the computation speed is also low [20]. Tao Peng et al. describe a method for segmentation of lung in chest radiographs with two main steps such as, image pre-processing method and refinement step to fine-tune the segmentation results [22].

Dr. Z. Faizal Khan et al. describe a method for automated segmentation of lung images using textural echo state neural networks. The lung nodules cannot be identified by using this method [6]. Sangheum Hwang et al. introduce an accurate lung segmentation model for chest radiographs. This method works based on deep convolutional neural networks [12]. Dhaval D. Kadia et al. describe a 3D Lung segmentation method which uses the deep learning model like U-Net. This work includes the application of medical imaging for diseases which seriously damages the lungs [14].

There are many existing methods involved in lung segmentation, but they are facing the limitations such as less segmentation-accuracy and high false segmentation. Major existing methods do not support left and right lung separation. The supervised learning based methods meet high complexity and high time consumption for lung segmentation. Hence, this research proposes a new algorithm for lung segmentation namely `Lung Segmentation based on Statistical analysis using Features of Connected Components (LS-SFCC)'. The proposed lung segmentation method segments the lungs using features of connected components. It performs the statistical data processing on features of foreground area and connected components. This method is enriched with morphological operations to gain high accuracy lung segmentation. It also supports an efficient left and right lung separation section. The highlight of this method is the light-weight statistical approach which produces an accurate lung segmentation with less complexity.

The Sect. 2 describes the working methodology of the proposed LS-SFCC method in a well organized manner. The Sect. 3 makes an deep analysis in lung segmentation algorithms using state-of-the-art analytic measures. The Sect. 4 makes a conclusion on the performance of the lung segmentation algorithms.

2 Proposed Methodology

The application of proposed work are:

- Used as a preprocessing application for lung cancer segmentation.
- To analyse the individual lung by the pulmonologists
- Can be developed as a application software for scanning machines
- Diagnosis of lung diseases such as lung cancer, pulmonary fibrosis, pneumonia and tuberculosis.
- Segmented images can be used as a key for treatment planning
- Used as an algorithm for computer-aided diagnosis of lung disease.

The novel research contributions involved in this research are: (a) statistical analysis based lung segmentation process and 0028b) texture descriptor based left and right lung separation process.

This statistical analysis based lung segmentation process works based on the well-designed decision-rules which are constructed by nature of different parts of lung CT image. The general strategy of dynamic-range-thresholds of lung-part and other-parts are carefully studied in the standard deviation based statistical analysis using the foreground area features and connected component features, from different lung images. Morphological operations involved in this work to accomplish the accurate lung segmentation. The dynamic range-threshold computation is the unique work of this research in lung segmentation which involves a brilliant ideology in standard deviation based histogram peak handling to construct the rule-set for lung segmentation.

The proposed method supports to extract the left and right lung regions separately in lung CT images. The Local binary pattern (LBP) is used to handle the lung image to have less complexity for the extraction of left and right lungs. Major existing methods have not support to the split of connected-lungs (left and right lungs are being in a connected-form), but the proposed method splits the lungs even though they are being in a connected form, which are called as left and right lungs. This method first computes the Local binary pattern computation for primary and secondary objects. After that, the connected state of lungs is detected using entropy, which is also a new approach in left and right lung separation process. Finally, the left and right lungs are isolated. A framework that contains the three concepts such as LBP, entropy and the simple splitting method, is developed to extract the left and right lungs with less-complexity and high accuracy.

The proposed method is robust to scaling issues. Lung images having various sizes can be easily adapted to this method without major issues due to the usage of statistical process. The medical industry consists two type of lung CT images: normal contrast lung image and high contrast lung image. This novel statistical based approach can be used as a generic method to segment both of these lung image types, because the statistical rule based system is carefully designed to meet this requirement.

The proposed method intakes the grayscale lung CT image and it produces three outputs viz. lungs segmentation output, left lung output and right lung output. This work can be branched into four divisions, and they are:

- Standard deviation based statistical data computation using foreground area feature
- Standard deviation based statistical data computation using connected component feature
- Lung segmentation based on peak analysis and morphological operations
- Left and right lung separation using LBP texture descriptor.

The first module computes the data analysis using foreground area feature. The second module computes the data analysis of connected component feature. The third module segments both lungs using histogram peak analysis and morphological operations. Finally, the left and right lung separation is accomplished based on LBP texture descriptor. These four modules are explained in detailed manner in the following sections.

The overall architecture of the proposed lung segmentation method is depicted in Fig. 1. This figure shows the internal blocks of foreground area feature computation, connected component feature computation, lung segmentation and left & right lung separation processes.
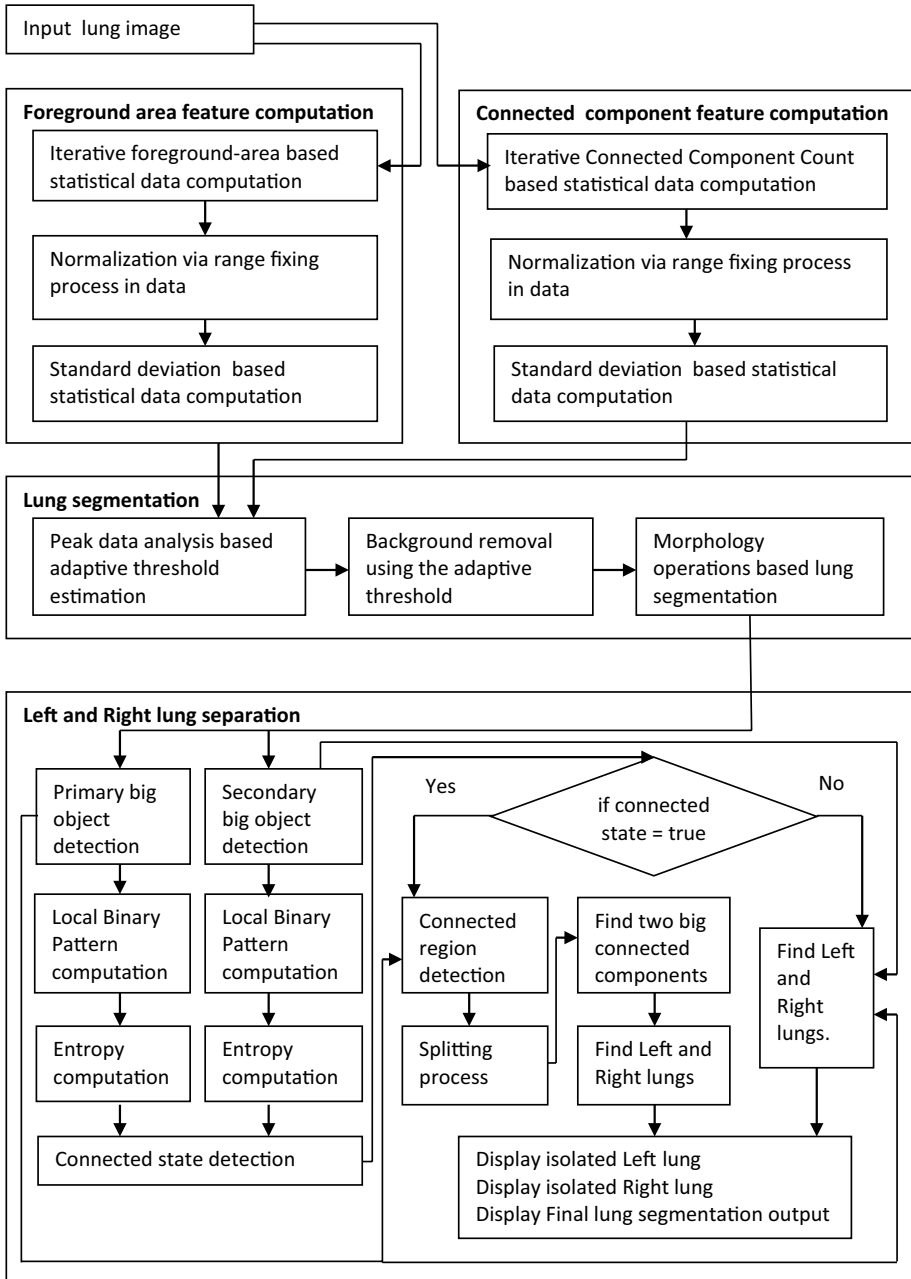
Fig. 1 Block diagram of the proposed LS-SFCC lung segmentation method

2.1 Standard Deviation Based Statistical Data Computation Using Foreground Area Feature

This section computes the standard deviation based statistical analysis with the help of area-property of foreground object. The aim of this section is to generate foreground's area-feature based histogram information which can be used as a feature data for lung segmentation section. The gray scale lung CT image of 512×512 dimension is given as the input to this research, and the output of this section is 'Area based histogram' and 'Standard deviation based histogram'. This work is sub divided into five small divisions and they are:

- Iterative binarization
- Area feature computation
- Area feature based histogram generation
- Normalization process
- Standard deviation feature based histogram generation.

2.1.1 Iterative Binarization

The gray scale lung image is constructed by 256 Gy shades. The gray value zero means highest dark intensity and the gray value 255 means highest white intensity. The gray value 128 means the combination of black and white. Among the intensities between 0 and 255, a specific intensity can be treated as an approximate threshold for lung segmentation. This approximate threshold can be found using an iterative approach by employing each intensity between 0 to 255 as a threshold to binarize the given input lung image. Binarization means grouping the image data into two clusters by making the data of image into two divisions which are formed by zero's and one's. The binarized output image contains only black and white pixels. This iterative binarization process provides an individual binarized output for each intensity. This process is explained in Eq. (1).

$$I_B^{i,j,k} = \begin{cases} 1, & if\ IL^{i,j} >= k \\ 0, & else \end{cases}$$
$$i \in [0, H-1], j \in [0, W-1], k \in [0, R] \tag{1}$$

Herein, the term I_B notifies the binarized image, IL indicates the input image, H specifies the image height, W describes the image width, k describes the intensity used for binarization, and R notifies the maximum intensity range. The term R possesses the maximum gray value of 255. The index variable k makes 256 binarizations using the input image.

Figure 2 illustrates the sample outputs of binarization process for the input lung image. Herein, only three binarization outputs, corresponding to the intensities/thresholds such as 64,128 and 192, are shown due to page limitation issues.

2.1.2 Area Feature Computation

In the binarized output, the pixels represented by 1 s are known as foreground pixels. The objects formed by 1 s are known as foreground objects. The area feature of the foreground is computed using Eq. (2).
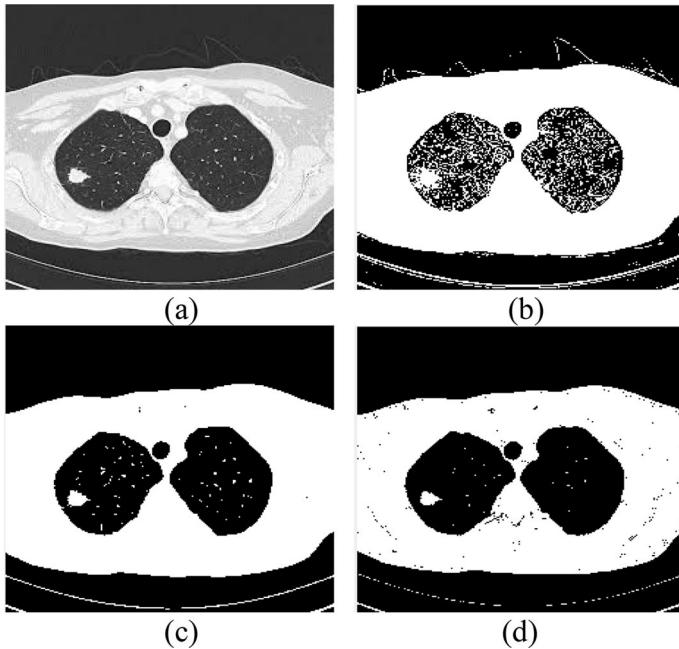
Fig. 2 Representation of binarization process, **a** input lung image, **b** binarized image for the threshold 64, **c** binarized image for the threshold 128, **d** binarized image for the threshold 192

$$F_A^k = \sum_{i=0}^{H} \sum_{j=0}^{w-1} I_B^{i,j,k}, \quad k \in [0, R] \tag{2}$$

In the Eq. (2), the term F_A^k refers the Area feature related to k^{th} threshold. The area feature is computed for each binarized image. The foreground pixel is constructed by the value 1. So, the summation of the binarized image yields the area feature related to the kth threshold.

2.1.3 Area Feature Based Histogram Generation

Histogram is a statistical data which shows the frequency of data occurrence. The area-feature F_A contains the area of 256 binarized images which are binarized by the intensity range of 0 to 255. The area feature describes the area of the foreground pixels which can be found through the count of 1 s. A histogram is a method of featuring an image using very limited range of array indexes. The 256 area features constitutes a 256 bin length histogram which represents the data that how many foreground area pixels appeared in each bin of histogram. This phenomena can be shown in Eq. (3).

$$H_A^k = F_A^k, \quad k \in [0, R] \tag{3}$$

Figure 3 expresses the histogram related to area feature. Herein, the 256 elements of this histogram are plotted as a chart.
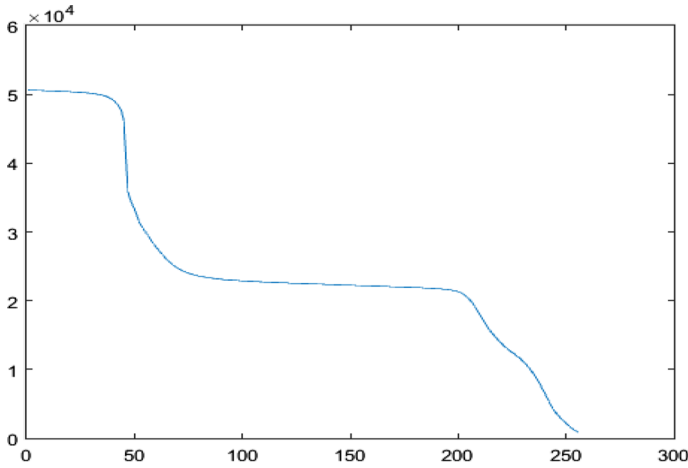
Fig. 3 Histogram of area-feature

2.2 Normalization Process

The shape of area feature histogram is varied for different lung images. Usually, the maximum peak value is not generated as unique, and it is not being in a standard form. A better analysis can be achieved, if the peak value is settled in a standard value such as 100. This work reforms the histogram so that the maximum peak value is reached by the standard value 100. This phenomena is called as normalization process. The Eqs. (4) and (5) illustrate the normalization process.

$$MX = Max(H_A) \tag{4}$$

$$H_{NA}^k = fix\left(\left(\frac{H_A^k}{MX}\right) * \propto\right), \quad k \in [0, R] \tag{5}$$

In Eq. (5), the term MX refers the maximum value of histogram H_A, Max refers the function to compute maximum value, fix refers the function to remove the fractional part, $\propto$ refers the normalization factor, and H_{NA} refers the normalized area feature based histogram. Herein, first the maximum value of H_A histogram is found using the $Max()$ function. After that each histogram element is divided by the maximum value MX, which yields the range of the value from 0 to 1. This result is multiplied by the normalization factor. Herein, normalization factor is fixed as hundred for an easy and effective analysis. Finally, the resultant values are fixed as integer using $fix()$ function. The normalized histogram of area feature is shown in Fig. 4.

2.2.1 Standard Deviation Feature Based Histogram Generation

Standard deviation is the measure of dispersion of a set of data from its mean. It computes the absolute variation of a distribution. If there is a higher variability then the standard
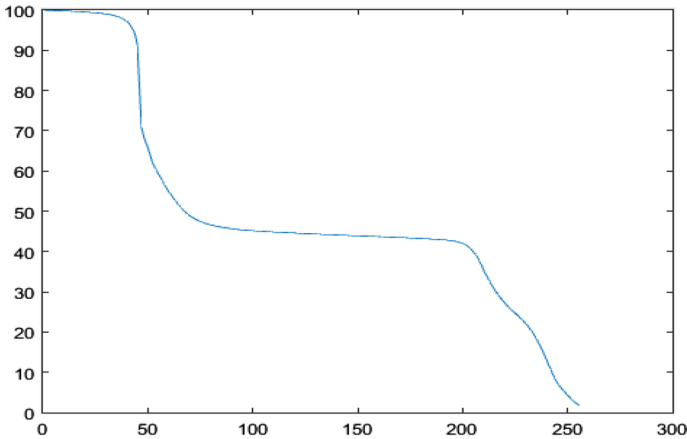
Fig. 4 Normalized histogram of area feature

deviation is also greater. Like that, if the variability is lesser, then the standard deviation is also lesser. Standard deviation points out the spread model of data. The steps of standard deviation computation are:

- Calculate the mean value
- Subtract each number from the mean value and square the resultant value
- Find the average of the squared differences computed by the previous step
- Compute the square root of the average value, which is known as the standard deviation value σ.

The aim of this section is to find out the flatten region in the area-histogram to find the better threshold. Normally, the threshold related to lung segmentation is existing in the region where the foreground area property is varied with less quantity. It can be computed using standard deviation histogram which is obtained by processing area-histogram. The standard deviation based histogram is computed using equations from Eqs. (6) to (11).

$$A = \{H_{NA}^{i-r}, H_{NA}^{i-2}, \cdots, H_{NA}^{i-r}\} \tag{6}$$

$$B = \{H_{NA}^{i+r}, H_{NA}^{i+2}, \cdots, H_{NA}^{i+r}\} \tag{7}$$

$$\sigma 1 = std(A) \tag{8}$$

$$\sigma 2 = std(B) \tag{9}$$

$$D = abs(\sigma 1 - \sigma 2) \tag{10}$$

$$H_{STDA}^{i} = D, \quad i \in [r, R - r] \tag{11}$$

Herein, the term A notifies the vector of left side elements in normalized area histogram, B refers the vector of right side elements in normalized area histogram, $\sigma 1$ refers

the Standard deviation of vector of left side elements, $\sigma2$ refers the standard deviation of vector of right side elements, *std* refers the function to compute standard deviation, *D* refers absolute difference of standard deviation of left and right side elements, *abs* refers the function to compute absolute value, H_{STDA} refers histogram of standard deviation, and *r* refers the region limit (Let it be 5). The Eq. (6) generates a vector having r elements which is constructed by the left side elements from the ith element of normalized area-histogram. Shortly speaking, the r elements along with left side of ith index is collected in the vector A. Herein, the region limit r is set by 5, which is derived from multiple trials of generation of highly informative histograms. This region limit r can be adaptive to major lung images. Hence, the vector A contains the left side 5 elements related to the ith index of histogram. The Eq. (7) generates the vector B which is constructed by the right side five elements. The Eq. (8) describes the computation of standard deviation of the left side elements. The Eq. (10) expresses the absolute difference computation of $\sigma1$ and 2. The term D describes that if the STD of left and right side elements is less than the ith index of area-feature property, it produces flatten dataflow in area histogram. Otherwise, it produces incremented or decremented dataflow in area-histogram.

The Eq. (11) creates the standard deviation histogram by assigning difference value D. In this histogram, the higher difference value produces peaks, and less difference value produces flatten surface.

Figure 5 depicts the shape of standard deviation based histogram. This histogram can be used to identify the locations of heavy modifications in the thresholding process, via peak data. The flatten regions of this histogram speaks about the less variation producing intensities in the thresholding process.

The standard deviation histogram has undergone the normalization process to make the histogram in the standard form having the highest data value of 100. This process is similar to the area-histogram normalization process, and it can be computed using Eqs. (12) and (13).
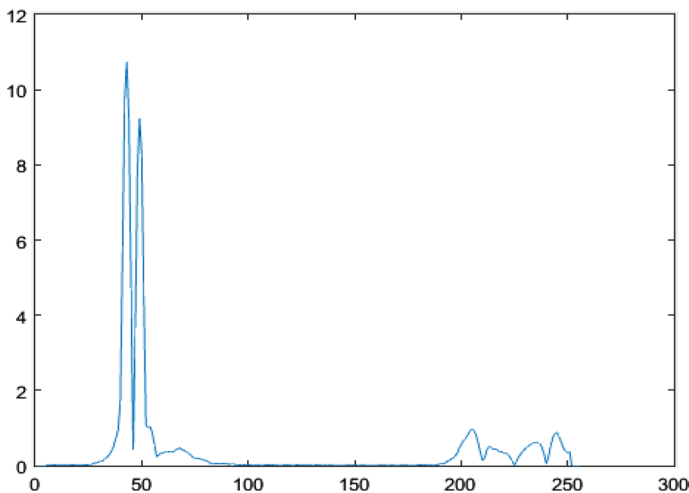

Fig. 5 Standard deviation based histogram

$$MX = Max\left(H_{STDA}\right) \tag{12}$$

$$H_{NSTDA}^k = fix\left(\left(\frac{H_{STDA}^k}{MX}\right) * \alpha\right) \tag{13}$$

The Eq. (12) computes the maximum value of STD-histogram. The Eq. (13) computes the normalized value of each element of STD-histogram using the parameters like H_{STDA}, MX and $\propto$. In the Fig. 6, the values of STD-histogram is bounded in the range of 0 to 100.

2.3 Standard Deviation Based Statistical Data Computation Using Connected Component Feature

In 2D image processing, connected components are defined as the clusters of pixels with the same value, that are connected to each other through either 4-pixel or 8-pixel connectivity. In case of 4-pixel connectivity, the clusters are formed by grouping the homogeneity pixels by contacting each other on either of their four faces. In 8-pixel connectivity, clusters of pixels are formed by connecting along any face or corner. In this section, the connected component count feature is used to segment the lung area. The aim of this section is to extract the connected component count features in the form of histogram which assists to a statistical analysis for lung segmentation. The doctrine of the usage of this feature is that the adaptive threshold for lung segmentation would lie in the intensity range where the less connected component count is generated. Shortly speaking, the intensity that involves with lower connecting components yields a better threshold for lung segmentation. Herein, 8-connectivity based connected component count feature is generated in the form of histogram. This work is sub divided into two parts and they are:

- Connected component feature based histogram generation
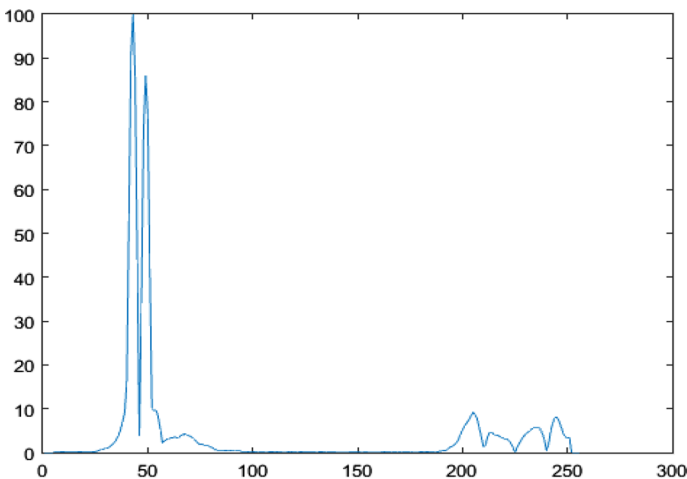- Standard deviation based connected component histogram generation.

Fig. 6 Normalized standard deviation based histogram

2.3.1 Connected Component Feature Based Histogram Generation

The iterative binarizations using the intensity range 0 to 255, are performed using input lung image via Eq. (1). Now, 256 binary images are obtained corresponding to 256 intensities. Then. the Connected Component Count (CCC) feature is computed. Finally, CCC-histogram and normalized CCC-histogram are generated. These processes are obtained by using equations from Eqs. (14) to (17).

$$F_{CCC}^k = Func\_CCC(I_B^{i,j,k}) \tag{14}$$

$$H_{CCC}^k = F_{CCC}^k \tag{15}$$

$$MX = Max(H_{CCC}) \tag{16}$$

$$H_{NCCC}^k = fix\left(\left(\frac{H_{CCC}^k}{MX}\right) * \propto\right) \tag{17}$$

$$k \in [0, R]$$

Herein, the term F_{CCC} refers the vector of connected component count feature, $Func\_CCC()$ refers the function to compute the CCC feature, H_{CCC} refers the histogram for connected component count, and H_{NCCC} refers the normalized histogram for connected component count feature. The Eq. (14) computes the connected component count feature from each binary image. The connected component can be imagined as the island like structures formed by 8-connectivity having the nature of possessing the binary values of 1 s. The count of the connected components in the kth binary image is stored in feature vector F_{CCC}. The CCC oriented is formed by Eq. (15). The normalization process is performed using Eq. (17) using the parameter α with the value of 100.

Figure 7 illustrates the normalized form of histogram related to CCC feature. This histogram shows the 'count' feature of connected component which occurred in each binary image of 256 intensities.
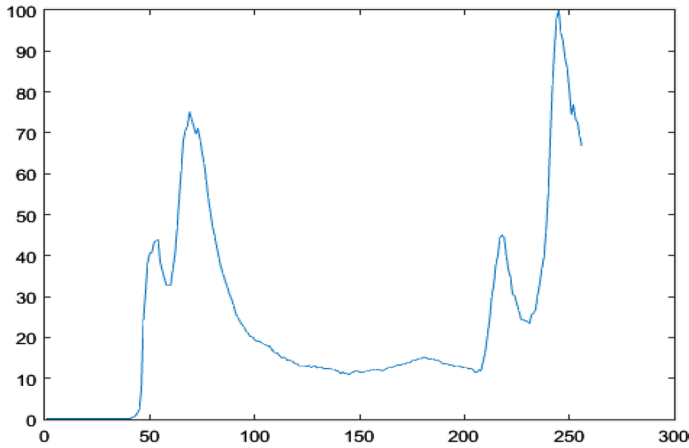
2.3.2 Standard Deviation Based Connected Component Histogram Generation

The standard deviation based CCC feature oriented histogram computation process is performed similar to the standard deviation based area feature computation model. Herein, the region limit is also preferred as 5. The left five elements of ith index of CCC-histogram, are involved to generate left side vector A based on Eq. (18). The right side CCC feature vector B is formed using Eq. (19).

$$A = \{H_{NCCC}^{i-1}, H_{NCCC}^{i-2}, \dots, H_{NCCC}^{i-r}\} \tag{18}$$

$$B = \{H_{NCCC}^{i+1}, H_{NCCC}^{i+2}, \dots, H_{NCCC}^{i+r}\} \tag{19}$$

The standard deviation value $\sigma 1$ is computed using the vector A. The standard deviation value $\sigma 2$ is computed using the vector B. The absolute difference D is computed using $\sigma 1$

Fig. 7 Normalized histogram for CCC feature

and $\sigma 2$ via Eq. (10). The histogram of standard deviation based connected component feature H_{STDC} is computed using Eq. (20).

$$H^i_{STDC} = D, \quad i \in [r, R - r] \tag{20}$$

This histogram has undergone the normalization process using Eqs. (21) and (22).

$$MX = Max(H_{STDC}) \tag{21}$$

$$H^k_{NSTDC} = fix\left(\left(\frac{H^k_{STDC}}{MX}\right) *\propto\right) \tag{22}$$

The Fig. 8 shows the normalized data of std based CCC feature histogram. It shows the deviation status of the CCC feature. This normalized histogram helps to detect the suitable threshold to segment lungs.

2.4 Lung Segmentation Based on Peak Analysis and Morphological Operations

This section segments the lung area of the query lung image using the standard deviation analysis of 'Area' and 'Connected Component Count' with the aid of morphological operations. The statistical analysis is performed using the histograms of both std-area feature and std-CCC features. The binarization threshold is computed using these features of histograms. Afterwards, morphological operations are applied to get the segmented lung area. This lung segmentation can be divided into four small blocks and they are: a) binarization via statistical based threshold, b) dominant object detection, c) morphological hole filling process, and image subtraction process.

### 2.4.1 Binarization Via Statistical Based threshold

There are two types of lung images available in the medical society. First one is normal contrast lung image and the second one is high contrast lung image. The normal contrast
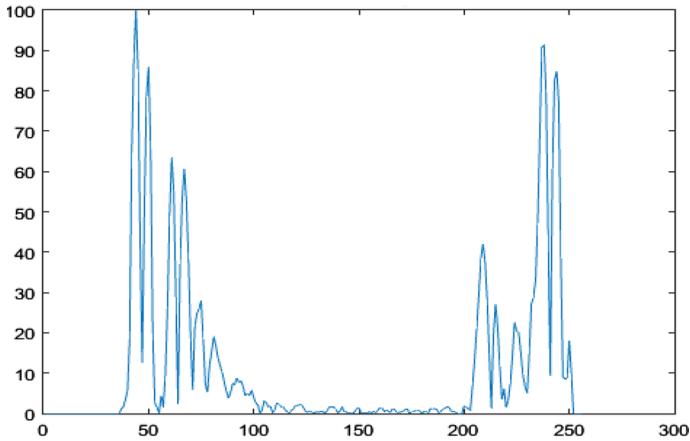
Fig. 8 Normalized histogram for std based CCC feature

lung image contains brighties foreground data, while the high contrast lung image contains non-brighties foreground data. In other words, normal contrast lung image contains less dark regions while the high contrast lung image is dominated by highly dark regions.

The Fig. 9 shows the aforementioned two types of lung images. The Fig. 9a shows the sample of normal contrast type image, and Fig. 9b illustrates the std oriented area histogram. Besides that, the Fig. 9c) depicts the std oriented CCC histogram. Figure 9d shows the sample of high contrast type lung image. In this image, the dark region is dominated.
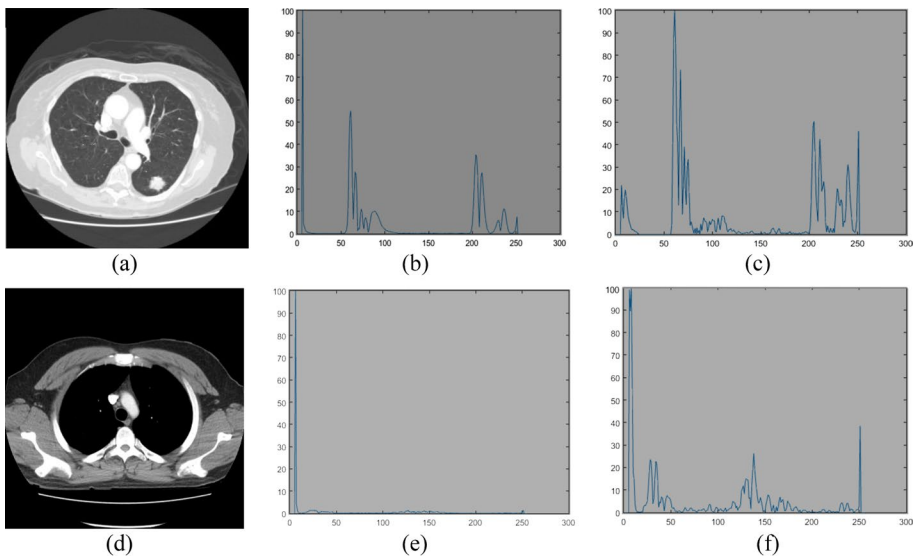
Fig. 9 Illustration of two types of lung images, **a** sample normal contrast type lung image namely sample-lung-1, **b** std-area histogram for sample-lung-1, **c** std-CCC histogram for sample-lung-1, **d** sample high contrast type lung image namely sample-lung-2, **e** std-area histogram for sample-lung-2 and **f** std-CCC histogram for sample-lung-2

Figure 9e shows the std oriented area histogram for the high contrast lung image. Figure 9f illustrates the std oriented connecting component count histogram. It can easily identify the contrast differences between Fig. 9a and d. The Fig. 9b contains the medium level difference between the first peak and the second peak. But Fig. 9e contains high range of difference between the first peak and second peak. The Fig. 9b and d shape a theory that the std-area histogram of normal contrast lung image yields less difference between first peak and second peak, meanwhile, the std-area histogram of high contrast lung image draws out high difference value between the first peak and the second peak. By making with lot of trials with many lung images, the minimum difference value can be set with 10% of the first peak. In other words, it can be spoken that if the difference between the first peak and the second peak is less than the 90% of the first peak, then the category of lung belongs to normal contrast model, otherwise, it is labelled to high contrast lung model. Figure 9f illustrates that the lung segmentation threshold can be fixed by the index value of first peak in case of high contrast lung images. The Fig. 9c depicts that the lung segmentation threshold can be defined as the middle value between the index of first-high peak and the index of next high individual peak, for the case of normal contrast lung image. Herein, a specific peak can be set as 'individual peak' with reference to first peak, when a touch occurs over the X-axis (or having zero y value). In other words, two peaks are considered as individual peaks, when there is at least a zero value i.e., related to Y-axis corresponding to the range of these two peaks. These innovations from the std-histograms of both the *area* and *CCC* features, segments the lung image into foreground and background objects. This phenomena can be explained using equations from Eqs. (23) to (29).

$$\beta = Find\ First\ Peak(H_{STDA}) \tag{23}$$

$$\gamma = Find\ Second\ Peak(H_{STDA}) \tag{24}$$

$$\delta = abs(\beta - \gamma \tag{25}$$

$$I_1 = f_1(H_{STDC}) \tag{26}$$

$$I_2 = f_2(H_{STDC}) \tag{27}$$

$$I_3 = f_3(H_{STDC}) \tag{28}$$

$$t = \begin{cases} I_1, & if\ \delta > (T_1 * \beta) \\ Min(I_1, I_2) + abs\left(\frac{I_1 - I_2}{2}\right), & if Validity(I_1 - I_2) = true \\ Min(I_1, I_3) + abs\left(\frac{I_1 - I_3}{2}\right), else \end{cases} \tag{29}$$

Herein, the term β refers the value of first high peak in std-area histogram, *FindFirst Peak*() refers the function to find the first high peak in the given histogram, γ refers the value of second high peak in std-area histogram, *FindSecondPeak*() refers the function to find the second high peak of given histogram, δ refers the absolute difference between the first high peak and the second high peak, I_1 refers the intensity corresponding to the first high peak in std-CCC histogram, $f_1()$ refers the function to find the intensity corresponding to first high peak in std-CCC histogram, I_2 refers the intensity corresponding to

the second high peak in std-CCC histogram, $f_2()$—refers the function to find the intensity corresponding to second high peak in std-CCC histogram, I_3 refers the intensity corresponding to the third high peak in std-CCC histogram, $f_3()$ refers the function to find the intensity corresponding to third high peak in std-CCC histogram, T_1 refers the threshold to determine the type of the lung image (let it be 0.9), t refers the threshold to segment the foreground and background object, and *FuncCheckValidity*() refers the function to check the validity of individual peak characteristics with reference to first high peak. The Eq. (23) finds the first high peak of H_{STDA} histogram. The Eq. (25) computes the absolute difference between the first high peak and second high peak values. The Eq. (26) determines the intensity corresponding to the first high peak in H_{STDC} histogram. The Eq. (27) determines the intensity corresponding to the second high peak of H_{STDC} histogram. The Eq. (28) computes the intensity corresponding to the third high peak of H_{STDC} histogram. The Eq. (29) computes the threshold t to segment the foreground and background objects. This process is performed using three steps and they are:

Step1: If the difference of first peak of std-area histogram and second high peak of std area histogram, is greater than the T_1 part of the first high peak, then the first high peak of H_{STDA} histogram is used as the threshold t.

Step2: If the condition $\delta > (T_1 * \beta)$ is false and the individual peak validity is true for I_2 with reference to I_1 then the middle value of I_1 and I_2 is used as the threshold t. In normal situations the term $abs\left(\frac{I_1 - I_2}{2}\right)$ is enough to compute the middle value. Suppose, the first peak originated after the second, then, the middle value computation must be used as $Min(I_1, I_2) + abs\left(\frac{I_1 - I_2}{2}\right)$.

Step3: If step1 and step2 is not eligible then step3 process occurs. Herein, the middle value of I_1 and I_3 is used as the threshold t.

In this way, the threshold for foreground and background object is computed. This threshold is used as a binarization threshold using Eq. (30) to get the Foreground object binary image

$$I_F^{i,j} = \begin{cases} 1, & if \ IL^{i,j} > t \\ 0, & else \end{cases} \qquad (30)$$
$$i \in [0, H-1], j \in [0, W-1]$$

The Fig. 10 illustrates the binary output after removing the background data. In this figure, the foreground objects are represented by the numeric value of 1.

2.4.2 Dominant Object Detection

The detection of highest area foreground object is the target of this section. The connected component concept is used to detect the dominant foreground object. The binary image I_F is taken as input and the connected components are detected based on 8 connectivity. Each connected component is labelled by 1,2,3 … n. The area of each connected component is computed by counting the pixels which are used to construct the connected component. The maximum area providing connected component is found, and that particular large connected component is called as the dominant foreground object.
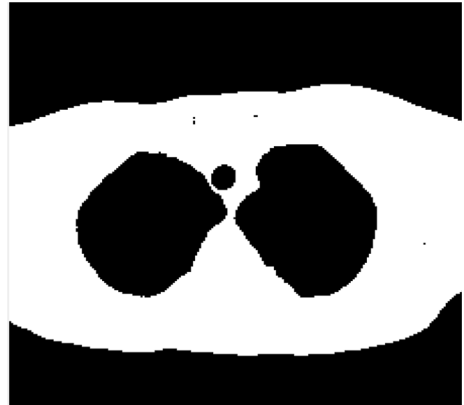
Figure 11 depicts the dominant foreground object of test Lung-1 image. In this image, the small foreground contents are removed and the highly dominant piece of foreground is chosen.

Fig. 10 Background removed image using the statistical analysis based threshold for Test-Lung-1 image
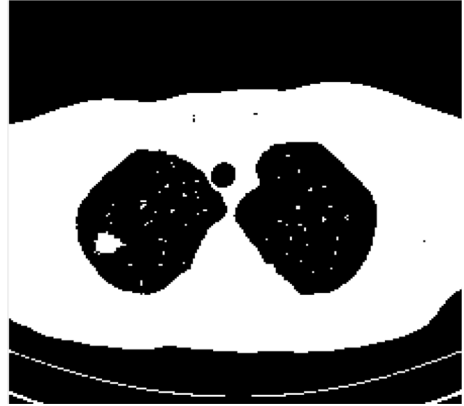
Fig. 11 Dominant foreground object of Test-Lung-1 image


2.4.3 Morphological Hole Filling Process

Normally, the dominant foreground object contains two holes corresponding to the left and right lung parts. These holes can be filled by using morphological operations. Herein, the term 'hole' means a set of background pixels that cannot be reached by filling the background from the edge of the image. The morphological operation namely dilation can be used to fill the holes of the dominant foreground object. The dilation fills the entire area of the hole in a specified set. The interaction at each set with the complement of the specified set limits the resultant area into the inside region of interest. In this way, the morphological process can be conditioned to meet a desired property. This process is known as 'conditional dilation'.

Figure 12 reveals the result of hole filling process in dominant foreground object based on conditional dilation process, which is a key process in morphological operations. The resultant output is stored as I_H.

Fig. 12 Hole filled output of
dominant foreground object for
Test-Lung-1 image
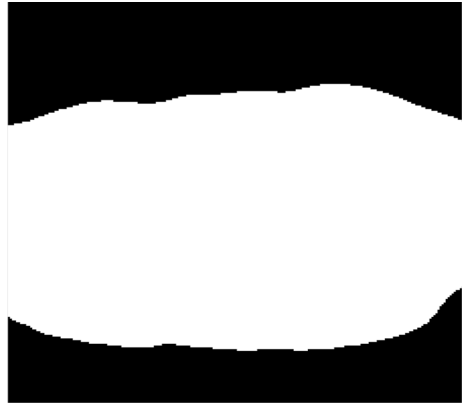
2.4.4 Subtraction Process

The hole filled output I_H and the dominant foreground object image I_F have undergone
the subtraction process to produce the binary lung image output. The same pixel locations
having different values, produce the lung segmented output as the foreground. This process
can be expressed as Eq. (31)

$$I_{BL}^{i,j} = abs(I_H - I_F), \quad i \in [0, H-1], j \in [0, W-1] \tag{31}$$

Herein, the term I_{BL} refers the Segmented binary lung image. Figure 13 reveals the lung
segmented output in binary form. The numeric value of lung object is set by 1 whereas the
background object is set by 0. There is an unwanted foreground object which is placed in
the resultant output. It should be removed.

2.5 Left and Right Lung Separation Using LBP Texture Descriptor

This process separates left and right lung individually. Suppose, the lungs are being in a con-
nected form then it would be split into two parts which are called as left and right lungs. The

Fig. 13 Binary lung segmented
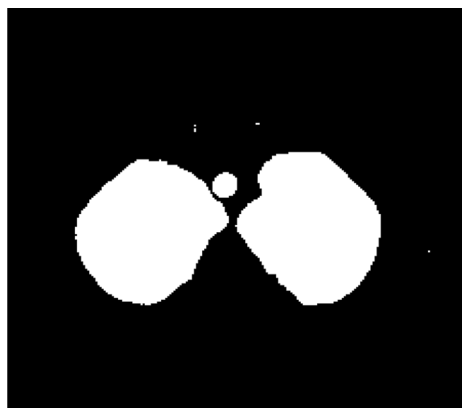output for Test-Lung-1 image

illustration of this process can be found in Fig. 1. The major steps of this novel method of left and right lung separation are:

- Local binary pattern computation for primary and secondary objects
- Connected state detection via entropy
- Splitting process for connected lungs
- Isolation of left and right lungs.

The biggest object is found in the partial-lung-segmented output I_{BL} via eight connected component concept and the resultant object is noted as primary big object. The second big object is found in the partial-lung-segmented output I_{BL} via eight connected component concept and the resultant object is noted as secondary-big object. Figure 14a illustrates the primary big object and Fig. 14b shows the secondary big object.

The Local Binary Pattern (LBP) process is applied over the primary object area by using the corresponding gray values of input image IL. The LBP is texture pattern which measures the textureness in the primary object area. The LBP texture image computation can be illustrated using Fig. 15 and equations from Eqs. (32) to (40).

$$B_0 = \begin{cases} 1, if(IL^{i,j} - IL^{i-1,j-1}) \geq 0 \\ 0, else \end{cases} \tag{32}$$

$$B_1 = \begin{cases} 1, if(IL^{i,j} - IL^{i-1,j}) \geq 0 \\ 0, else \end{cases} \tag{33}$$

$$B_2 = \begin{cases} 1, if(IL^{i,j} - IL^{i-1,j+1}) \geq 0 \\ 0, else \end{cases} \tag{34}$$

$$B_3 = \begin{cases} 1, if(IL^{i,j} - IL^{i,j+1}) \geq 0 \\ 0, else \end{cases} \tag{35}$$

(a)   (b)
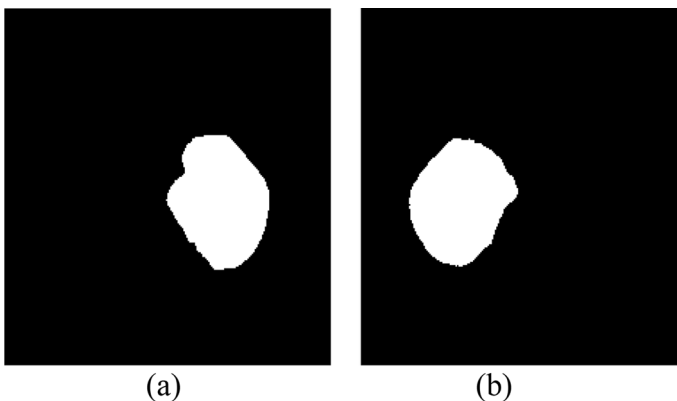
Fig. 14 Primary and secondary objects representation output for Test-Lung-1 image, **a** primary big object and **b** secondary big object

Fig. 15 Representation of LBP computation, **a** neighbour window, **b** order of LBP computation and **c** weight values used in LBP

$IL^{i-1,j-1}$	$IL^{i-1,j}$	$IL^{i-1,j+1}$
$IL^{i,j-1}$	$IL^{i,j}$	$IL^{i,j+1}$
$IL^{i+1,j-1}$	$IL^{i+1,j}$	$IL^{i+1,j+1}$

(a)

0	1	2
7		3
6	5	4

(b)

2^0	2^1	2^2
2^7		2^3
2^6	2^5	2^4

(c)

$$B_4 = \begin{cases} 1, if(IL^{i,j} - IL^{i-1,j-1}) \geq 0 \\ 0, else \end{cases} \tag{36}$$

$$B_5 = \begin{cases} 1, if(IL^{i,j} - IL^{i+1,j}) \geq 0 \\ 0, else \end{cases} \tag{37}$$

$$B_6 = \begin{cases} 1, if(IL^{i,j} - IL^{i+1,j-1}) \geq 0 \\ 0, else \end{cases} \tag{38}$$

$$B_7 = \begin{cases} 1, if(IL^{i,j} - IL^{i,j-1}) \geq 0 \\ 0, else \end{cases} \tag{39}$$

$$\begin{aligned} I_{LBP1}^{i,j} = & (B_7 * 2^7) + (B_6 * 2^6) + (B_5 * 2^5) + (B_4 * 2^4) \\ & + (B_3 * 2^3) + (B_2 * 2^2) + (B_1 * 2^1) + (B_0 * 2^0) \\ & i \in [0, H-1], j \in [0, w-1] \end{aligned} \tag{40}$$

Figure 15a illustrates the neighbouring window centred by the pixel $IL^{i,j}$. The Fig. 15b shows the order of the processing of LBP, and Fig. 15c shows the weight values corresponding to the neighbour positions. Herein, the term B_0 refers the 0th bit of LBP, B_1 refers the 1st bit of LBP, …, and B_7 refers the 7th bit of LBP. The term I_{LBP1} refers the primary object LBP image. The first bit of LBP is computed using Eq. (32). If the difference of centre pixel $NF^{i,j}$ and 0th order neighbour pixel, is positive, then the first LBP bit B_0 is set by 1, which is otherwise set by 0. In this way, the other bits from B_1 to B_7 obtain their bit values as either 1 or 0 using corresponding equations. The primary object LBP image is computed using Eq. (40) by the summation process of multiplication of bit values and the corresponding weight values. The second big object has also undergone the LBP process

similar to the primary object, and the resultant image is stored in I_{LBP2}. The Fig. 16a shows the LBP image for primary object and Fig. 16b denotes the LBP image for secondary object.

Generally, the entropy can be defined as the statistical measure of randomness that can be used to characterise the texture of the input image. The entropy measurement scheme is well explained in [8]. The entropy measurement is defined as $\sum p * log_2(p)$. Herein, p contains the histogram counts of input image. The I_{LBP1} image has undergone the entropy process, and the output value is stored as e_1. The I_{LBP2} image has also undergone the entropy process, and the output is stored as e_2.

The connected lungs property can be effectively computed using Eq. (41). Herein, the term C_p refers the connection property value, and T_e refers the threshold entropy.

$$C_p = \begin{cases} 1, if \left(\frac{e_2}{e_1} \right) < T_e \\ 0, else \end{cases} \tag{41}$$

In Eq. (41), the connection property value is set by 1, if it has a connected lung structure, which is otherwise set by 0. Generally, the disconnected form of lungs generate the ratio of the entropy values of the secondary and primary objects is closer to 1. Also, the connected form of lungs generates the ratio of $\left(\frac{e_2}{e_1} \right)$ to less than 0.5. In case of connected lung format, the primary object is constructed by the connected lungs and the secondary object is constructed by the tissue-objects. The threshold T_e is set by 0.5 in Eq. (41). Always the entropy of lung area is higher than the tissue area, so the heavy variation in the entropy ratio of primary and secondary objects reveal the connected lung scenario in lung images. The less variation in the entropy ratio of primary and secondary object reveals the disconnected form of lung image. The threshold T_e is set by 0.5 by making trials in hundred test images.

Suppose, both the lungs are in connected form then it should be split into two parts, i.e., left lung and right lung. The narrow connected-column in the primary object is detected via the column-wise histogram. The column-wise histogram is generated based on primary object using $\frac{1}{4}th$ of column value to $\frac{3}{4}th$ of column value. In the resultant column wise histogram, the narrow connected column Ω which is represented by the lower value of histogram, is found. In the binary primary object image, the entire binary values of that specific

(a)   (b)
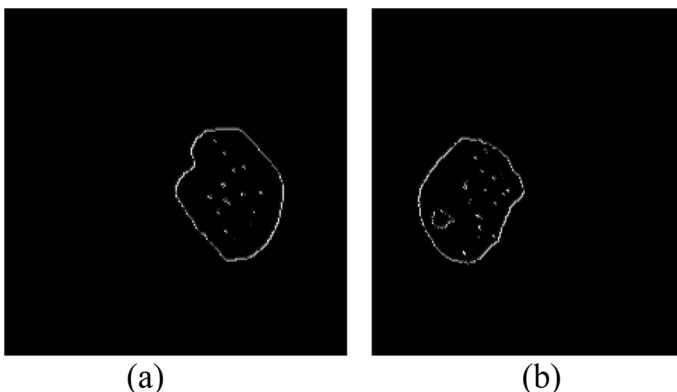
Fig. 16 LBP results, **a** LBP image for primary object, **b** LBP image for secondary object

narrow connected column are replaced by the values of 0 s. This phenomenon splits the region of connected lungs into left and right lung objects. The object data at the left side of Ω is separated as *Binary left lung image*. The object data at the right side of Ω is separated as *Binary right lung image*. The binary form of left lung image is noted as I_{BLL} and the binary form of right lung part is noted as I_{BRL}. The binary image which contains both the lung image in the separated-arrangement is noted as I_{BSL} which mentions the *Binary segmented lung image*.

Suppose, the lungs are not set with connected form by representing $C_p = 0$, then the left and right lung objects are detected individually via the closest column value of zero. The left-most column value of the primary object is found and marked as C_1, and the left most column value of secondary object is found and noted as C_2. If $C_1 < C_2$ is true, then the primary object is noticed as left lung object I_{BLL}, otherwise, the secondary object is noted as left lung image I_{BLL}. If $C_1 > C_2$ is true, then the primary object is noticed as the right lung image I_{BRL}, otherwise, the secondary object is noted as the right lung image I_{BRL}. The binary image which contains both the lung image in the disconnected form is quoted as I_{BSL} which means the binary segmentation lung image.

The region of left lung is filled using the real data from the input image *IL*, and the resultant gray output is quoted as I_{GLL}. The real gray lung data from the input image *IL* is projected over the binary right lung region of the image I_{BRL}, and the gray output is noted as I_{GRL}. The real gray values noted as from the image *IL* are projected over the foreground region of I_{BSL} and the resultant gray image is marked as I_{GSL}, that means the Gray-segmented-lung image.

The Fig. 17a shows the Binary left image segmentation output. Figure 17b shows the Binary right image segmentation output. The Fig. 17c shows the binary segmentation image for both lungs.

The Fig. 18 shows the segmentation results for the Test-Lung-1 image in gray format. Figure 18a shows the left lung segmentation output in gray format while the Fig. 18b shows the right lung segmentation output in gray format. The Fig. 18c shows the segmentation output related to both the left and right lungs.

The outputs such as Binary left lung image I_{BLL}, Binary right lung I_{BRL}, Binary both lung image I_{BSL}, Gray left lung image I_{GLL}, Gray right lung image I_{GRL}, and Gray both lung image I_{GSL} are showcased to the user. Thus, the lung region segmentation process is performed.

3 Discussion and Analysis

In this section, the proposed LS-SFCC method is analysed against three benchmarked databases, and they are:

- LIDC database [17]
- LCTSC database [16]
- KGMC database [13].

The *Lung Image Database Consortium (LIDC)* image collection include the lung cancer screening thoracic Computed Tomography (CT) scans with marked-up annotated lesions [17]. This database possesses 1018 cases of lung cancer diagnosis. The proposed research uses 250 lung CT images from the LIDC database as training data sets. This dataset is
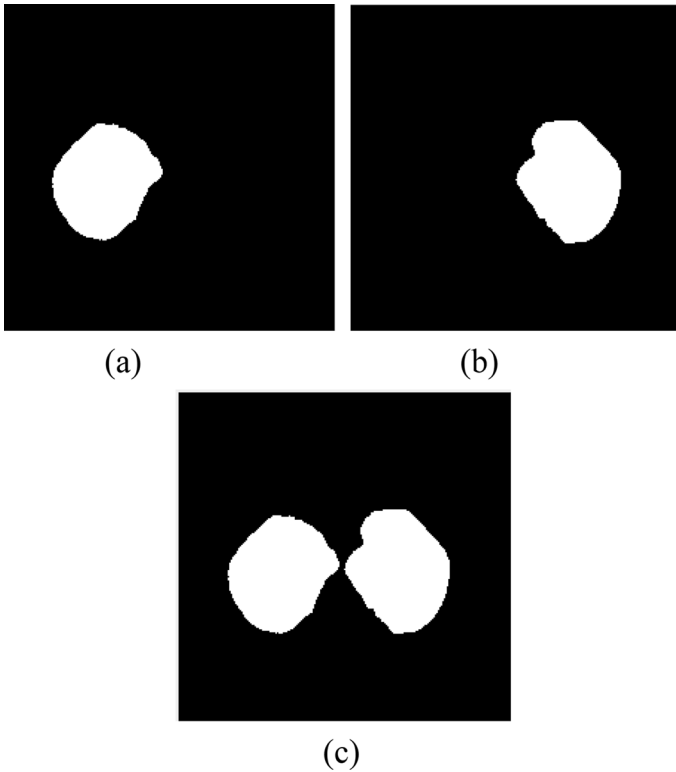
Fig. 17 Segmentation outputs in binary form for Test-Lung-1 image, **a** left lung image I_{BLL}, **b** right lung image I_{BRL}, **c** both left and right lungs image I_{BSL}

annotated through this thesis as *LIDC-DB*. The Fig. 19 depicts the sample images of the LIDC-DB database.

The *Lung CT Segmentation Challenge* 2017 (LCTSC) is associated with the challenge competition and conference session held on 2017 [16]. The LCTSC database consists of 9593 lung images to support the lung diagnostic oriented researches. The proposed research of this thesis chooses 250 lung images as test data from the LCTSC database. This test database is spelled throughout this thesis by the term *LCTSC-DB*. The Fig. 20 depicts the sample images of the LCTSC-DB database.

The *Kanyakumari Government Medical College (KGMC)* is a multi speciality hospital [13]. Earlier, this hospital was known as Govt. T.B Hospital. The Government Medical College was started in the year 2004. The department of Pulmonology has progressed in an effective manner. The proposed research of this thesis is involved with 250 lung CT images which are received from this hospital. This dataset is termed as *KGMC-DB* database throughout this thesis. The Fig. 21 depicts the sample images of the KGMC-DB database.

The comparison part of the proposed LS-SFCC method is organized using the three recent existing methods such as:

- Lung Segmentation using MLevelSet method (LS-MLS) [1]
- Lung Segmentation using U-Net semantic segmentation method (LS-UNET) [19]
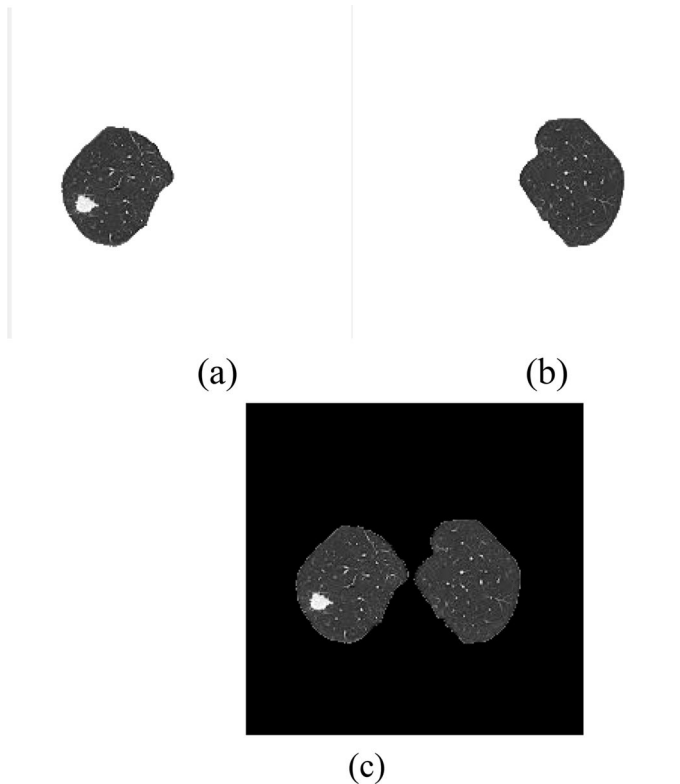
Fig. 18 Segmentation outputs in gray form for Test-Lung-1 image: **a** left lung image I_{GLL}, **b** right lung image I_{GRL}, **c** both left and right lungs image I_{GSL}

- Lung segmentation using Color based Fuzzy C Means clustering method (LS-CFCM) [15].

The existing methods such as LS-MLS and LS-CFCM and the proposed method's performance evaluation is found using the entire 250 images per database as test images. The existing LS-UNET method is implemented by taking the 125 images from each databases as training images and another 12 images as test images.

Figure 22 depicts the sample screenshots of the proposed LS-SFCC method. Herein, Fig. 22a shows the input lung image, Fig. 22b shows the normalized area histogram, Fig. 22c displays the normalized std-area histogram, Fig. 22d shows the normalized connected component histogram, Fig. 22e shows the normalized std-connected component histogram, Fig. 22f describes the background removed lung image, Fig. 22g shows the hole filled image, Fig. 22h shows the intermediate segmented binary lung image, Fig. 22i depicts the primary big object, Fig. 22j shows the secondary big object, Fig. 22k shows the LBP of primary big object, Fig. 22l describes the LBP of secondary big object, Fig. 22m focuses the lung image after splitting, Fig. 22n depicts the binary segmented left lung image, Fig. 22o illustrates the binary segmented right lung image, Fig. 22p shows the binary segmented output for left and right lungs, Fig. 22q points out the segmented left lung image with projected gray values, Fig. 22r showcases the segmented right lung image
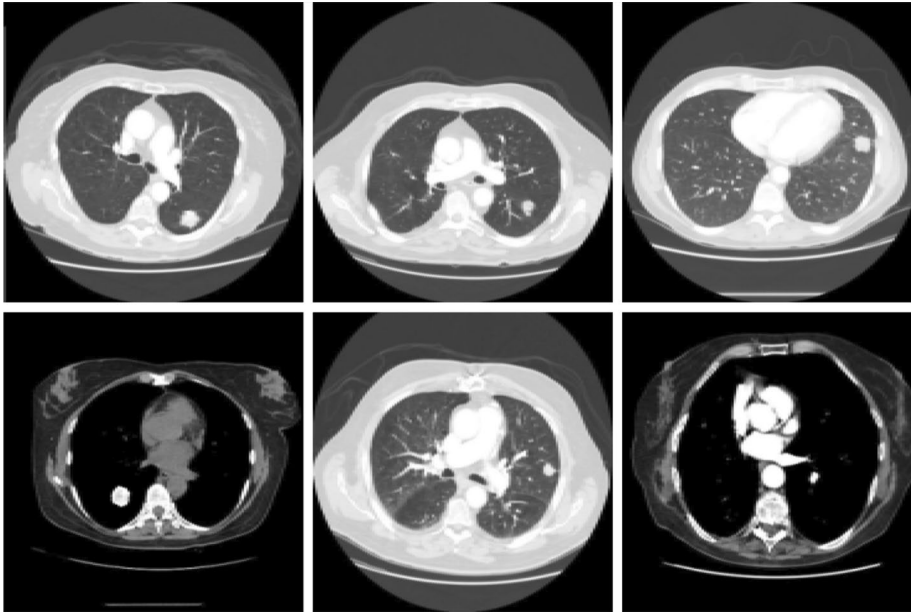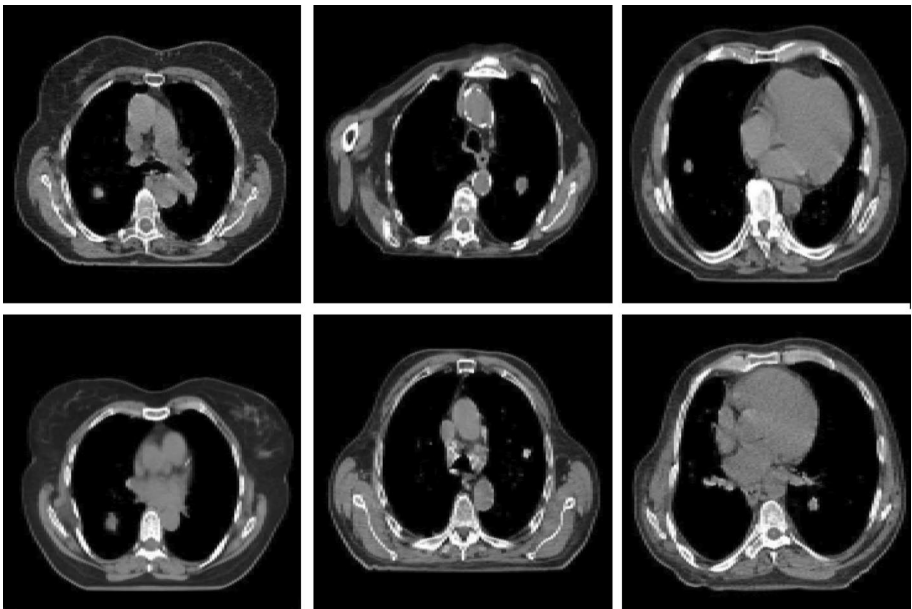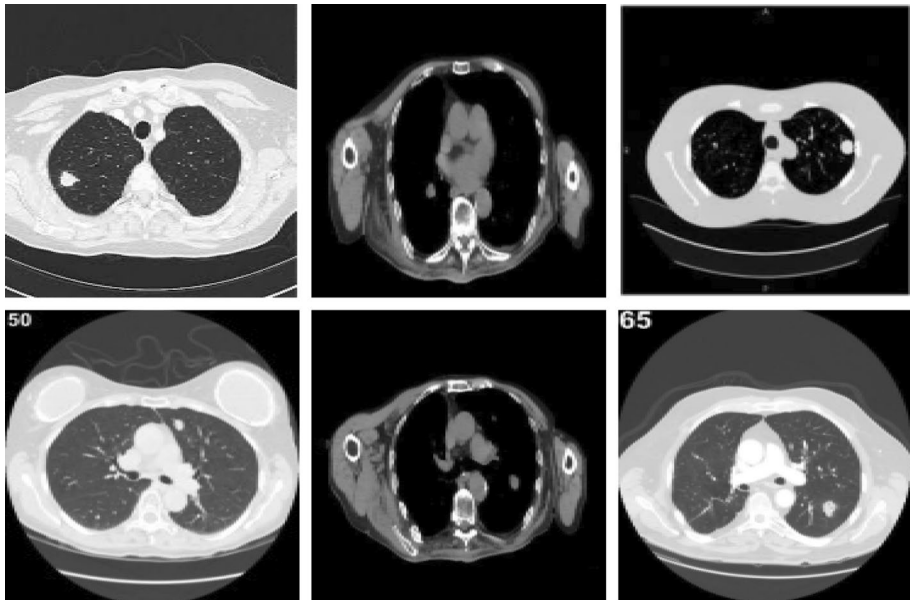
Fig. 19 Sample images of LIDC-DB database

Fig. 20 Sample images of LCTSC-DB database

Fig. 21 Sample images of KGMC-DB database

with projected gray values and Fig. 22s shows the segmented output for left and right lungs with projected gray values.

Segmentation-MSE is evaluated by the similarity between the segmented-image and the Ground-truth image. If the MSE value is less, then the quality of lung segmentation is high. If the MSE_{LS} value is high, then quality of segmentation is low. Table 1 shows the MSE analysis of nine lung images. The lung images are taken from LIDC-DB database, LCTSC-DB database and KGMC-DB database.

The minimum MSE value refers to the best lung segmentation method. The lowest MSE provider of this analysis is the proposed LS-SFCC method for lung segmentation. The lower value of the MSE of the LS-SFCC method for lung segmentation is 0.0505, which is corresponding to the LCTSC-DB-1 image. The average MSE is computed for 100 test images from each database. The resultant average MSE of lung segmentation for LIDC-DB is 0.0738. The resultant average MSE of lung segmentation for LCTSC-DB is 0.0673. The resultant average MSE of lung segmentation for KGMC-DB is 0.0839.

Table 1 describes the MSE analysis for lung segmentation. Generally, less MSE indicates the best segmentation because MSE is an 'error' based analytic metric. The least MSE corresponding to LIDC-D6 is provided by the proposed LS-SFCC method. Also, it yields the least MSE values corresponding to the other databases such as LCTSC-DB and KGMC-DB. The least MSE values generated by the proposed methods are 0.0603, 0.0505 and 0.0752 with respect to the LIDC-DB, LCTSC-DB and KGMC-DB databases. But, the second best method LS-CFCM generates higher MSE values than the proposed method and less MSE values than the other two existing methods such as LS-MLS and LS-UNET. The average MSE of the proposed method is 0.075 whereas it is 0.1875 for the next-best method. Hence, the proposed method improves the MSE by 60% when compared to the next-best existing method. Since the proposed method yields the least MSE value for the three databases, it is noted as the best lung segmentation method.
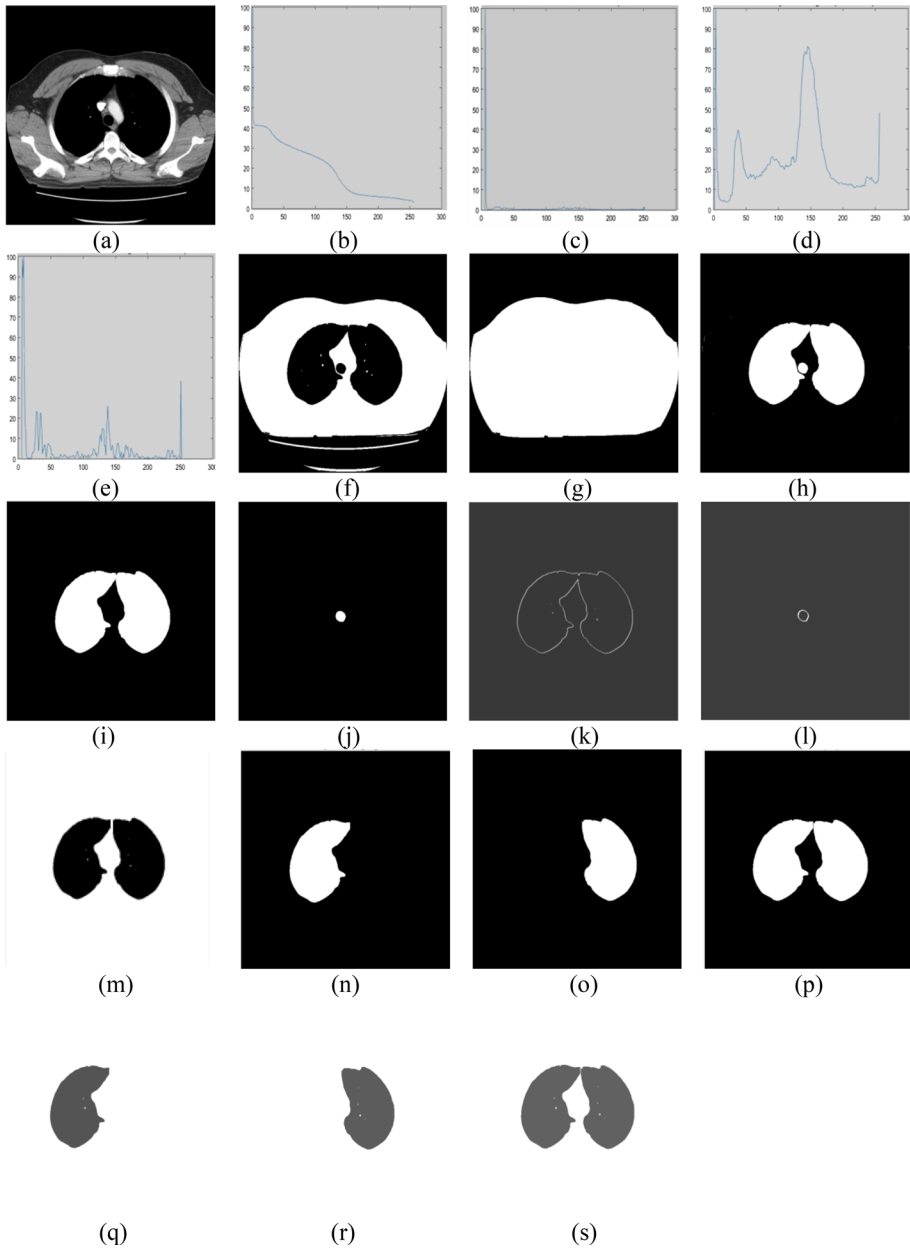
Fig. 22 Sample outputs for the proposed lung segmentation method

The PSNR computes the peak Signal-to-noise ratio between two images. It is measured in decibels. The segmentation-PSNR measures the similarity between the Lung segmented image and Ground-truth image. The lung segmentation-PSNR $PSNR_{LS}$ is computed using Eq. (42).

Table 1 MSE analysis for lung segmentation

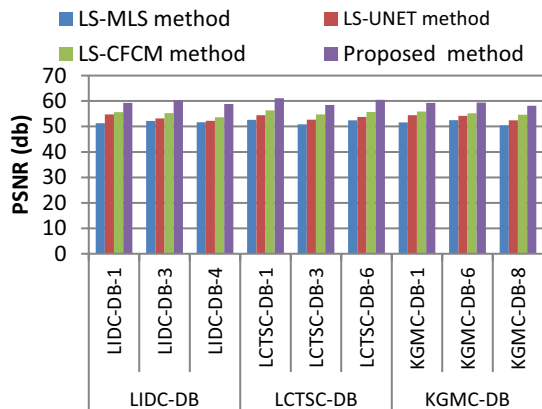| Image	MSE of lung segmentation			
	LS-MLS method	LS-UNET method	LS-CFCM method	Proposed LS-SFCC
LIDC-DB-1	0.4810	0.2174	0.1783	0.0764
LIDC-DB-3	0.3928	0.3149	0.1933	0.0603
LIDC-DB-4	0.4407	0.3910	0.2839	0.0846
LCTSC-DB-1	0.3549	0.2351	0.1511	0.0505
LCTSC-DB-3	0.5347	0.3541	0.2189	0.0932
LCTSC-DB-6	0.3699	0.2743	0.1747	0.0581
KGMC-DB-1	0.4478	0.2356	0.1683	0.0766
KGMC-DB-6	0.3657	0.2478	0.1964	0.0752
KGMC-DB-8	0.5917	0.3751	0.2234	0.1001

$$PSNR_{LS} = 10 * log_{10}\left(\frac{255^2}{MSE}\right) \tag{42}$$

If $PSNR_{LS}$ value is high then the corresponding method is considered as the best lung segmentation method. If a method's $PSNR_{LS}$ value is low then, it is the poor lung segmentation method. Figure 23 shows the PSNR analysis of nine lung images.

According to the theory, higher PSNR means the better segmentation method. The highest PSNR provider of this analysis is the proposed LS-SFCC method for lung segmentation. The Fig. 23 shows the PSNR assessment for lung segmentation related to various lung databases. The highest value of the PSNR of the LS-SFCC method is 61.098 db, which is corresponding to the LCTSC-DB-1 image. The average PSNR is computed for 100 test images from each databases. The resultant average PSNR are 59.495 db, 60.008 db, 58.928 db corresponding to the databases LIDC-DB, LCTSC-DB and KGMC-DB respectively.

Segmentation Accuracy is used to make assessment about the performance of the proposed LS-SFCC method against the existing methods. It is evaluated using Eq. (43). The unit of segmentation accuracy measurement is percentage (%).


Fig. 23 Chart for PSNR assessment for lung segmentation related to various lung databases

Table 2 Segmentation accuracy analysis for lung segmentation

Image	Segmentation accuracy (%)			
	LS-MLS method	LS-UNET method	LS-CFCM method	Proposed method
LIDC-DB-1	87.70	88.47	90.94	94.74
LIDC-DB-3	88.33	90.05	92.27	95.55
LIDC-DB-4	86.75	87.27	90.49	93.63
LCTSC-DB-1	87.61	88.36	92.53	96.38
LCTSC-DB-3	88.13	88.94	91.25	94.41
LCTSC-DB-6	87.90	89.44	91.98	95.37
KGMC-DB-1	88.10	89.11	90.04	94.58
KGMC-DB-6	87.31	88.05	89.72	94.42
KGMC-DB-8	87.60	88.79	90.05	93.66

Fig. 24 Chart for segmentation accuracy assessment for lung segmentation related to lung database
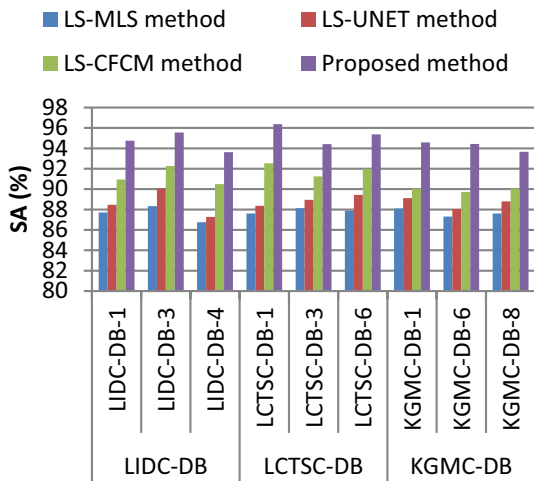
$$SA_{LS} = \frac{TP + TN}{TP + FP + TN + FN} \tag{43}$$

In Eq. (44), The term TP (True Positive) refers to the number of pixels that are truly segmented with reference to the target object whereas the term FP (False Positive) refers to the number of pixels that are falsely segmented with reference to the same target objects. The term TN (True Negative) denotes the number of pixels that are truly segmented with reference to background segmentation whereas the term FN (False Negative) denotes the number of pixels that are falsely segmented with reference to the background segmentation. Segmentation accuracy gives the performance quality of the particular segmentation method. If the segmentation accuracy is higher, then, the corresponding segmentation method is considered the best segmentation method and vice versa. The maximum SA value refers to the best lung segmentation method. Table 2 and Fig. 24 show the segmentation accuracy analysis of nine lung images.

The Table 2 describes the segmentation accuracy analysis for lung segmentation. The proposed method achieves the highest segmentation accuracy values for the three databases compared to the existing methods. The highest value corresponding to LIDC-DB database is 95.55% which is generated for the LIDC-DB-3 image. The highest segmentation accuracy value corresponding to the LCTSC-DB database is 96.385% which is yielded by the LCTSC-DB-1 image. The higher segmentation accuracy value corresponding to the KGMC-DB database is 94.52% which is given by the KGMC-DB-1 image. The high segmentation accuracy of the proposed method makes it the best one, when compared to the other three existing methods. The LS-CFCM method provides the next-best values compared to the other two existing methods, so it is noted as the next-best method in lung segmentation in terms of segmentation accuracy. The resultant average segmentation accuracy for the proposed LS-SFCC method are 94.64, 95.38 and 94.22% with respect to the LIDC-DB, LCTSC-DB and KGMC-DB databases. Meanwhile the resultant average segmentation accuracy for LS-CFCM method are 91.23, 91.92 and 89.93% with respect to the LIDC-DB, LCTSC-DB and KGMC-DB databases. These values prove that the performance level of the proposed method is the best one when compared to the existing methods. The overall segmentation accuracy for the proposed method and LS-CFCM method are 94.74 and 91.02% respectively. So the proposed method improves the segmentation accuracy than the next-best method by 4.08% which proves the potential performance of the proposed lung segmentation method.

The highest SA provider of this analysis is the proposed LS-SFCC method. The lowest SA provider of this analysis is the LS-MLS method for lung segmentation. The higher value of the SA of the LS-SFCC method for lung segmentation is 96.38%, which is corresponding to the LCTSC-DB-1 image. The average SA is computed for 100 test images from each database. The resultant average SA using the proposed LS-SFCC method for LIDC-DB is 94.64%. The resultant average SA using the proposed LS-SFCC method for LCTSC-DB is 95.38%, and for KGMC-DB is 94.22%.

The EPQI-LSM analysis examines the segmented-output lung images through eye perception, and it provides an index value based on that the quality of the segmented method's grade. This index value is known as Eye perception based quality index. The quality analysis is performed based on only human eye perception. The EPQI value reflects the performance grade of the segmentation methods. If the index value is higher, then the concerned segmentation-method's performance is best one and vice-versa. Figure 25 shows the Eye perception based quality index for lung segmentation methods. The three databases such as

Fig. 25 Chart for Eye Perception based quality index for lung segmentation methods
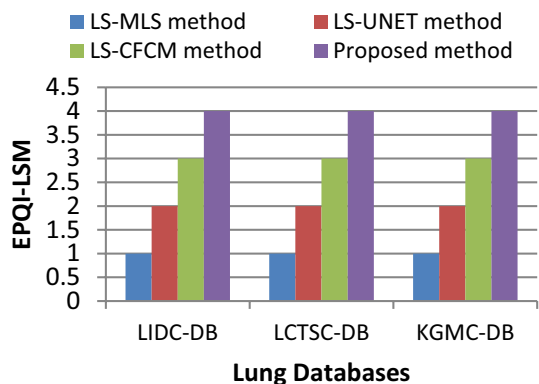
Fig. 26 Chart for time-taken for lung segmentation methods
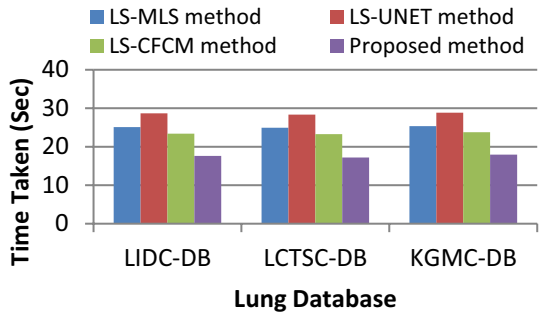
Table 3 EPQI-LD analysis for lung segmentation

Method	EPQI-LD		
	LIDC-DB	LCTSC-DB	KGMC-DB
LS-MLS	2	3	1
LS-UNET	2	3	1
LS-CFCM	2	3	1
Proposed LS-SFCC	2	3	1

LIDC-DB, LCTSC-DB and KGMC-DB are considered to progress the assessment of lung segmentation.

This analysis is done by ten human observers who visually examined the output of 100 test images from each database by eye perception. The human observers rank each method, according to the output quality, by following the guideline that better method should be projected by higher index rank and vice versa. In Fig. 25 the proposed LS-SFCC method successfully reaches the higher rank i.e., fourth rank, which makes it a better method than the others, for the three databases. The second best method which can be indicated by the index value 3 is LS-CFCM method. The least rank holder is the LS-MLS method that holds the EPQI value as 1.

Time-taken analysis of lung segmentation algorithms helps to assess the time–cost efficiency of various algorithms of lung segmentation. Figure 26 shows the time-taken analysis of lung segmentation methods. The Fig. 26 shows the time-taken for various lung segmentation methods.

From this assessment, the proposed LS-SFCC method produces good quality in lung segmented image with lowest time taken value. The lowest time taken for lung segmentation is occupied by LS-SFCC method, which is 17.20 s for the LCTSC-DB.

The EPQI-LD analysis is performed based on the eye perception only. Table 3 shows the EPQI-LD analysis for the three databases such as LIDC-DB, LCTSC-DB and KGMC-DB regarding the lung segmentation performance. These three databases are considered to progress the assessment of lung segmentation through the four methods such as LS-MLS, LS-UNET, LS-CFCM and the proposed LS-SFCC. This analysis is done by ten human observers who visually examined the output of 100 test images from each database by eye perception. The highest rank is given for LCTSC-DB database. The lowest rank is given to KGMC-DB database.

F-Score analysis measures the segmentation quality through two parameters namely Precision and Recall which are computed based on Eq. (44).

Table 4 Average F-score analysis for lung segmentation

Method	Average F score
LS-MLS	0.9063
LS-UNET	0.9181
LS-CFCM	0.9202
Proposed LS-SFCC	0.9421

$$F - Score = 2 * \left(\frac{Precision \times Recall}{Precision + Recall} \right) \tag{44}$$

Precision is computed using the two parameters *True Positive* (TP) and *False Positive* (FP). Recall is computed using two parameters *True Positive* (TP) and *False Negative* (FN). In common, the higher F-Score means better segmentation quality. The lesser F-Score means fair segmentation quality.

In this assessment, 150 test images are chosen from the three concerned databases in the composition of $50 + 50 + 50$. The FScore results are computed for these images, and the average FScore is computed. Table 4 expresses the Average F-Score values of the four methods on lung segmentation. The proposed LS-SFCC method provides the average FScore value as 0.9421. The LS-CFCM method generates the average FScore value as 0.9202. The average FScore analysis decides that the proposed LS-SFCC method is the best method in lung segmentation than the existing methods, because it holds with highest FScore value than other methods.

4 Conclusion

The proposed LS-SFCC method segments the lung region in lung CT images effectively via three approaches: (a) Standard deviation based statistical data computation using foreground area feature and connected component feature, (b) Lung segmentation based on peak analysis and morphological operations, (c) Left and right lung separation using LBP texture descriptor. These approaches enhances the LS-SFCC method to robust against scaling problem. The characteristics of lung regions in CT images are deeply studied through more than 500 lung images from three databases, and based on that study, the statistical based decisions are designed. The proposed method achieves the best results such as 61.098 db, 96.38% and 0.9421 corresponding to peak signal to noise ratio, segmentation accuracy and FScore. None of the methods provide exact segmentation results with the ground-truth image, but it can be claimed that the proposed method gives a better approximation than the other methods. The LS-SFCC method segments the lung regions with high speed execution. Hence, the proposed LS-SFCC method is considered to be chief lung segmentation method than the existing methods. This segmentation method is designed to apply only on lung CT images. In future, this research can be extended to lung segmentation in MRI images also.

Declarations

Conflict of interest The authors do not have conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent All methods were carried out in accordance with relevant guidelines and regulations.

Human and Animal Rights Humans and animals are not involved in this research.

References

1. Chae, S. H., Moon, H. M., Chung, Y., Shin, J. H., & Pan, S. B. (2016). Automatic lung segmentation for large-scale medical image management. *Springer, Multimedia Tools Application, 75*, 15347–15363.
2. Cheimariotis, G.-A., Al-Mashat, M., Haris, K., Aletras, A. H., Jogi, J., Bajc, M., Maglaveras, N., & Heiberg, E. (2017). Automatic lung segmentation in functional SPECT images using active shape models. *Springer, 32*(2), 94–104.
3. Chen, G., Xiang, D., Zhang, B., Tian, H., Yang, X., Shi, F., Zhu, W., Tian, B., & Chen, X. (2019). Automatic pathological lung segmentation in low dose CT image using eigenspace sparse shape composition. *IEEE Transactions on Medical Imaging, 38*(7), 1736–1749.
4. Chung, H., Ko, H., Jeon, S. J., Yoon, K. H., & Lee, J. (2018). Automatic lung segmentation with Juxta-Pleural nodule identification using active contour model and bayesian approach. *IEEE Journal of Translational Engineering in Health and Medicine, 6*, 1–13.
5. Dai, S., Ke, Lu., Dong, J., Zhang, Y., & Chen, Y. (2015). A novel approach of lung segmentation on chest CT images using graph cuts. *Elsevier, 168*, 799–807.
6. Faizal khan, Z., Al Sayyari, A. S. & Quadri, S. U. (2017). Automated segmentation of lung images using textural echo state neural networks. In: IEEE, Confernce 2017.
7. Filhoa, P. P. R., Cortez, P. C., Da Silva Barrosc, A. C., Albuquerque, V. H. C., & Tavares, J. M. R. S. (2017). Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images. *Elsevier Medical Image Analysis, 35*, 503–516.
8. Gonzalez, R. C., Woods, R. E., & Edlis, S. L. (2003). *Digital image processing using MATLAB*. Prentice Hall. chapter 11.
9. Guol, S., & Wang, L ., (2015). Automatic CT image segmentation of the lungs with an iterative Chan-Vese algorithm. In: *IEEE, Conference*, 2015.
10. Hadavi, N., Nordin, M. J., & Shojaeipour, A. (2014). Lung cancer diagnosis using CT-scan images based on cellular learning automata. *In: 2014 International Conference on Computer and Information Sciences (ICCOINS),* Kuala Lumpur, pp. 1–5.
11. Huidrom, R., Chanu Y. J., & Singh, K. M. (2017). A fast automated lung segmentation method for the diagnosis of lung cancer. In: IEEE Conference, 2017.
12. Hwang, S., & Park, S. (2017). *Accurate lung segmentation via network-wise training of convolutional networks* (pp. 92–99). Springer.
13. KGMC database, Available from: www.KKMC.ac.in/kkmc/index.jsp. Accessed on [4 Feb 2020].
14. Kadia, D. D., Alom, Z., Burad, R., Nguyen, T. V., & Asari, V. K. (2021). R 2U3D: Recurrent residual 3D U-Net for lung segmentation. *IEEE Access, 9*, 88835–88843.
15. Khan, Z. F. (2019). Automated segmentation of lung parenchyma using colour based fuzzy C-Means clustering. *Springer, 14*, 2163–2169.
16. LCTSC database, Available from: https://wiki.cancerimagingarchive.net/display/Public/Lung+CT+Segmentation+Challenge+2017. Accessed on [3 Feb 2020].
17. LIDC database, Available from: https://wiki.cancerimagingarchive.net/display/Public /LIDC-IDRI#. Accessed on [2 Feb 2020].
18. Mansoor, A., Bagci, U., Foster, B., Ziyue, Xu., Papadakis, G. Z., Folio, L. R., Udupa, J. K., & Mollura, D. J. (2015). Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends. *Radio Graphics, 35*(4), 1056–1076.
19. Nemoto, T., Futakami, N., Yagi, M., Kumabe, A., Takeda, A., Kunieda, E., & Shigematsu, N. (2019). "Efficacy evaluation of 2D, 3D U-Net semantic segmentation and atlas-based segmentation of normal lungs excluding the trachea and main bronchi" IEEE indexed journal. *Journal of Radiation Research, 61*(1), 257–264.
20. Nurfauzi, R., Nugroho, H. A., & Ardiyanto, I., (2017). Lung detection using adaptive border correction.In: IEEE Conference 2017.

21. Oluyide, O. M., Tapamo, J.-R., & Viriri, S. (2018). Automatic lung segmentation based on graph cut using a distance-constrained energy. *IET Computer Vision, 12*(5), 609–615.

22. Peng, T., Wang, Y., Xu, T. C., & Chen, X. (2019). Segmentation of lung in chest radiographs using hull and closed polygonal line method. *IEEE Access, 7*, 137794–137810.

23. Peng, T., Xu, T. C., Wang, Y., Zhou, H., Candemir, S., Mimi, W., Zaki, D. W., Ruan, S. J., Wang, J., & Chen, X. (2020). Hybrid automatic lung segmentation on chest CT scans. *IEEE Access, 8*, 73293–73306.

24. Pinzon, A. M., Orkisz, M., Richard, J. C., & Hoyos, M. H. (2017). Lung segmentation by cascade registration. *Elsevier, 38*(5), 266–280.

25. Soliman, A., Khalifa, F., Elnakib, A., El-Ghar, M. A., Dunlap, N., Wang, B., Gimel'farb, G., Keynton, R., & El-Baz, A. (2017). Accurate lungs segmentation on CT chest images by adaptive appearance-guided shape modeling. *IEEE Transactions on Biomedical Engineering, 36*(1), 263–276.

26. Zhang, W., Wang, X., Zhang, P., & Chen, J. (2017). Global optimal hybrid geometric active contour for automated lung segmentation on CT images. *Computers in Biology and Medicine, 91*, 168–180.

27. Zhou, J., Yan, Z., Lasio, G., Huang, J., Zhang, B., Sharma, N., Prado, K., & D'Souza, K. W. (2015). Automated compromised right lung segmentation method using a robust atlas-based active volume model. *Elsevier, 46*(Part 1), 47–55.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

V. Juliet Rani received her MCA degree from Noorul Islam College of Engineering, Kumaracoil, Thucklay, Tamil Nadu in 2003. She received M.Phil in Computer Science from Mother Teresa university,Kodaikanal, India in 2004. She is a research scholar of Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India. She has 16 years of teaching experience. At present she is working in St. Jerome's College, Nagercoil, Tamil Nadu, India. Her research area is Medical Image Processing.

Dr. K. K. Thanammal is currently working as an Associate Professor in the Department of Computer Science and Centre for Research at S.T. Hindu College, Nagercoil, Tamil Nadu, India. She received Ph.D(computer science) & M.Phil(computer science) from Manonmaniam Sundaranar University, Tirunelveli, India, M.Sc(computer science) from Bharathidasan University, Thiruchirapalli, India & B.Sc (computer science) from Madurai Kamaraj University, Madurai, India. She passed SLET(2016) conducted by Mother Teresa University, Kodaikanal, India. She has 30 years of teaching experience. She published 25 journals in various National and International journals. She is a research Supervisor in Manonmaniam Sundaranar University. Her main areas of interest are Medical Image Processing, Artificial Intelligence, Data Mining and Neural Networks. She is acting as a board chairman for BCA and MCA board of studies in Manonmaniam Sundaranar University,Tirunelveli, India, since 2015 onwards.