International Journal of Soft Computing 10 (2): 127-136, 2015
ISSN: 1816-9503

Two-Phase Supervised Segmentation Algorithm for Automatic Segmentation of Lung Parenchyma from Chest CT

C. Sunil Retmin Raj, H. Khanna Nehemiah, D. Shiloah Elizabeth and A. Kannan
Anna University, 600025 Chennai, Tamilnadu, India

Abstract: Segmentation of lung parenchyma is a challenging task in the Computer Aided Diagnosis (CAD) of lung disorders using chest Computed Tomography (CT). In this research, a Two Phase Supervised algorithm has been proposed for segmentation of lungs in chest CT slices. In the first phase, the initial lung region is obtained by applying a combination of iterative thresholding and morphological operations. The shape features of the resulting lung region are applied to a decision tree classifier that is constructed from a training dataset to determine whether the segmented lung forms a complete lung. In the second phase, if the initial lung is complete the lung region is filled with lung tissue if the initial lung is not complete, the lung region is determined by a series of operations. First, the longest of the two connected components is determined. The longest connected component is then folded and translated horizontally. The two lung regions are then converted to a single connected component and the convex hull is obtained. The convex hull is interpolated to obtain the outer convex edge. The outer convex edge thus obtained is superimposed on the binary image obtained by folding and translation and used as the initial contour for the Active Contour Model (ACM). The ACM algorithm is iterated until the distance between the contours of two subsequent iterations becomes lesser than a threshold. It is also ensured that the number of components does not exceed two. This method is adaptive in that the number of iterations of ACM is not fixed and is based on the image for which it is applied. This method of lung segmentation has been compared with the conventional Iterative Thresholding Method, Convex Hull Based algorithm and Supervised algorithm for segmentation. The maximum overlap achieved with all the four methods is 100% while the minimum achieved with the proposed method is 55.3%, conventional iterative thresholding method is 37.83%, Convex Hull Based algorithm is 25.82% and Supervised algorithm is 54.25%. Thus, the proposed Two-Phase Supervised Method is found to be better than the other three methods with which the comparison is done.

Key words: Lung, computer aided diagnosis, segmentation, active contour model, convex hull

INTRODUCTION

Lung disorders include numerous diseases and are a major concern to the society. They can be treated if diagnosed early. Nodules cannot be classified based on location alone. One of the most reliable methods for distinction between malignant nodule and benign nodule is the growth rate (Collins *et al*., 1956, Nathan *et al*., 1962). Such measurements can be done by computer Aided Diagnosis (CAD) Systems only when lung segmentation and subsequent Region of Interest (ROI) extraction techniques are automated. Nodules can be isolated or attached to the chest wall.

Segmentation of lung in the presence of solitary lung nodules has been well studied and there are numerous techniques for extraction of such nodules. Juxta-pleural nodules are the ones attached to the lung wall. Automated segmentation of lung in the presence of such nodules is a challenging task. The challenge lies in automatically identifying the boundary between the chest wall and the nodule. Research is still in progress in the field of segmentation of lung from chest CT that bears large and dense peripherally placed nodules.

Hu in their research have proposed a fully automatic method for identifying the lungs in 3D x-ray CT images. They have achieved it in three steps. First, gray level thresholding has been applied to extract the lungs. Then, the lung border has been identified using dynamic programming and finally, a set of morphological operations have been applied to smoothen the lung border. They have tested their proposed method by processing 3D CT datasets of eight normal subjects. The average root mean square difference between the manually segmented lung and the lung segmented by the proposed method has been found to be 0.58 pixels.

Corresponding Author: C. Sunil Retmin Raj, Anna University, 600025 Chennai, Tamilnadu, India

Elizabeth *et al.* (2012a, b) have proposed an Automated Method for segmentation of lung parenchyma from chest CT. They have achieved it in two steps. First the lungs have been segmented using optimal thresholding and then the lung border has been reconstructed using the convex hull and centroid properties. They have tested their method by using it in a CAD System for diagnosis of lung cancer and have achieved an accuracy of 97% as compared to 88.5% achieved using a segmentation based on optimal thresholding. The CAD System using the proposed method for segmentation has been found to detect large peripherally placed Pathology Bearing Regions (PBRs) that couldn't be detected by the CAD System that uses optimal thresholding based segmentation but still was found to work badly in the presence of very large peripheral PBRs.

Darmanayagam *et al.* (2013) have proposed a supervised approach for segmentation of lung from chest CT. This involves training a BPN with shape feature to distinguish between complete and incomplete lung and using it to determine if the lung extracted by thresholding and morphological operations is complete. If the outcome of this initial segmentation is not complete the complete lung is determined, reflected and translated to get the other lung. It has been shown that the approach is capable of segmenting the lung completely even in the presence of larger peripherally placed PBRs provided both the lungs are not severely affected. They have tested their approach in a CAD System for diagnosis of bronchiectasis, tuberculosis and pneumonia and have achieved an accuracy of 97.37%.

Wei *et al.* (2013) have proposed an automatic method for segmenting and repairing the lung parenchyma. They have combined the optimal iterative threshold, three-dimensional connectivity labeling, three-dimensional region growing for the initial segmentation of the lung parenchyma and an improved chain code and Bresenham algorithms to repair the lung parenchyma. They have tested their approach on 122 chest CT scans. They have compared their research with the most cited rolling ball operator (Armato and Sensakovic, 2004). They have achieved 95.2% of segmentation accuracy of lung parenchyma and 100% juxtapleural nodule inclusion sensitivity.

To summarize, iterative thresholding is simple and efficient when the PBRs are placed internally but not efficient in the presence of peripherally placed PBR. Convex Hull Based algorithm handles this issue in cases where the peripherally placed PBR is of moderate size but not efficient in the presence of large peripherally placed PBRs. ACM is capable of handling large PBRs at the peripherally but the segmentation result depends on the number of iterations which is not determined adaptively. The supervised approach reconstructs the lung with large peripherally placed PBRs but may include few non-lung regions and may not be efficient ehen both the lungs are severely affected.

Compared to the researches discussed in the literature, the research presented in this study differs in the following ways: it combines the advantages of iterative thresholding (Elizabeth *et al.*, 2009), Convex Hull Based algorithm (Elizabeth *et al.*, 2012a, b), ACM and supervised approach (Darmanayagam *et al.*, 2013) for segmentation of lung from chest CT eliminating their drawbacks. In addition the optimal set of features for determining the completeness of the lung segmented by the conventional thresholding and morphological operations has been determined experimentally by constructing decision trees using different sets of features.

SUPERVISED TWO PHASE SEGMENTATION ALGORITHM

Segmentation of lungs using the proposed approach takes place in two phases, namely, determination of the completeness of lung and extraction of lung parenchyma as shown in Fig. 1. The steps involved in each phase is shown in detail in Fig. 2.

Dataset generation: A set of CT chest image slices are segmented using the Conventional Thresholding Based algorithm. The shape features of the two connected components that represent the lung regions are extracted and a feature vector consisting of the ratios of width,

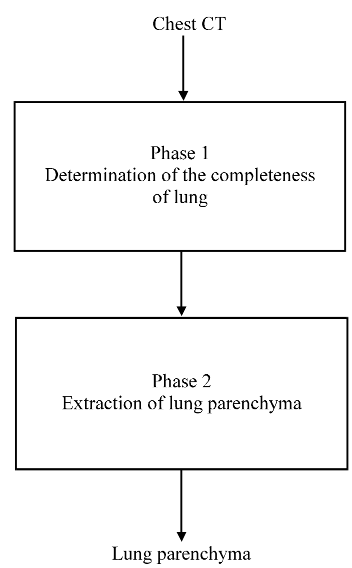

Fig. 1: Phases of the Two-Phase Supervised Lung Segmentation algorithm

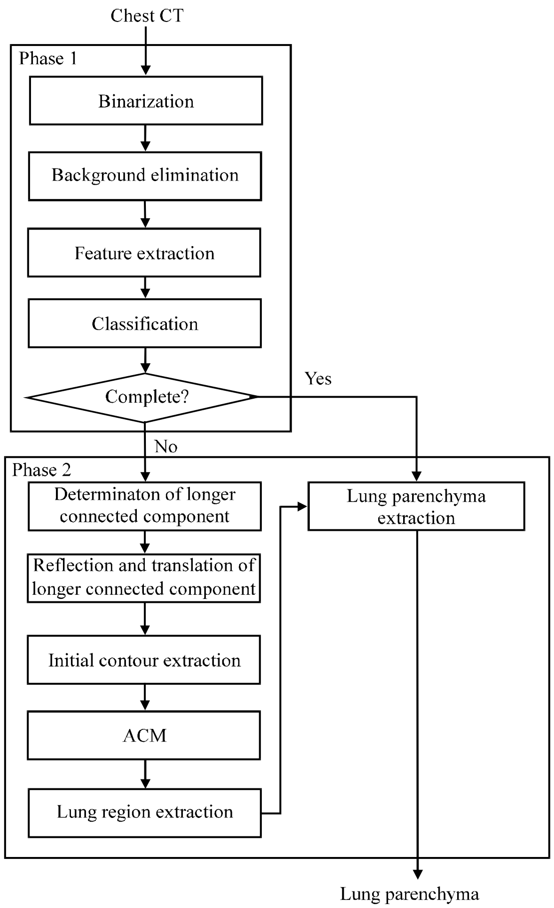

Fig. 2: Steps involved in the Two Phase Lung Segmentation algorithm

height, area, convexity and eccentricity of the left lung region to that of the right lung region and the convexity, width, height, area and eccentricity of the individual lungs is constructed. The feature set and the corresponding class labels are used to construct a decision tree using CART algorithm (Breiman *et al*., 1984):

Input:
Training set comprising of chest CT slices
Process logic:
Step 1: The image is converted to binary image by applying iterative thresholding (Ridler and Calvard, 1978).
a. The global threshold (t) of the image is determined (Otsu, 1979).
b. The initial threshold t_0 is set to t.
c. The image, f(x,y) is converted to binary image, b(x,y) by Eq. 1:

$$b(x,y)=\begin{cases}1, & \text{if } f(x,y)\geq t\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

d. The average intensity, μ_1 of pixels in f(x, y) that are set to 1 in b(x, y) and the average intensity, μ_0 of pixels in f(x, y) that are set to 0 in b(x,y) are determined.
e. The new threshold, t is the average of μ_1 and μ_0 as defined by Eq. 2:

$$t=\frac{\mu_1+\mu_0}{2} \tag{2}$$

f. Steps 1c-1e are repeated until the thresholds of the current and previous iterations are equal.
Step 2: Morphological operations are applied and background is removed.
a. Non-lung components with an eccentricity >0.95 are removed (Experimentation showed that regions with an eccentricity >0.95 cannot be lungs)
b. The image obtained in step 2 a is complemented.
c. The white pixels not connected to the boundary are set to black.
Step 3: The airways are removed by setting the black pixels that are not connected to the boundary to white.
Step 4: The shape features of the two connected components that represent the lung regions in the binary image, namely, convexity, area, eccentricity, equiv diameter, height, width, perimeter, major axis length, minor axis length are computed and a feature vector consisting of the ratios of width, height, area, convexity and eccentricity of the left lung region to that of the right lung region and the convexity, width, height, area and eccentricity of the individual lungs is constructed.
Step 5: The feature vector is labeled manually as 2 if the segmented lung is complete and 1 if it is incomplete (Labeling has been done manually as this is the training phase)
Step 6: Steps 1 through 5 are repeated for each chest CT slice in the training set.
Output:
Dataset consisting of the feature vectors generated in steps 1 through 6 and the corresponding class labels assigned in step 5

Construction of decision tree classifier: A decision tree is constructed using the dataset by applying CART algorithm (Breiman *et al*., 1984) that uses Gini index as a measure of impurity.

Input:
Dataset, D, consisting of feature vectors of size 15 belonging to classes 1 and 2
Process logic:
Step 1: A root node is created for the tree.
Step 2: The feature, A that best classifies the dataset is determined.
a. The prior probabilities, p_i that a feature vector in the dataset D belongs to class C_i is estimated using Eq. 3:

$$p_i=\frac{|C_{i,D}|}{|D|} \tag{3}$$

where, $|C_{i,D}|$ is the number of feature vectors that belong to class Ci and |D| is the number of feature vectors in the dataset.
b. The Gini index of the set of training feature vectors in dataset, D is computed as defined by Eq. 4:

$$\text{Gini}(D)=1-\sum_{i=1}^{m}p_i^2 \tag{4}$$

where, m is the number of classes.
c. For each feature A,
i. For each possible binary split find the Gini index using Eq. 5:

$$\text{Gini}_A=\frac{|D_1|}{D}\text{Gini}(D_1)+\frac{|D_2|}{|D|}\text{Gini}(D_2) \tag{5}$$

where, D_1 and D_2 are the partitions of the dataset.

ii. The binary split giving the minimum $Gini_A$ is chosen as the split point for that attribute.
iii. The reduction in impurity that would be incurred by the split on feature A is computed using Eq. 6:

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (6)$$

iv. The feature that maximizes the reduction in impurity, $\Delta Gini(A)$ is selected as the splitting attribute used in the node.

Step 3: For each possible split, s_i of A, a new branch is added below the root node
  a. If all the feature vectors in the dataset belong to class 1, the node is labeled 1.
  b. If all the feature vectors in the dataset belong to class 2, the node is labeled 2.
  c. If the set of attributes is empty, the node is labeled with the class to which majority of the feature vectors belong.
  d. If there are feature vectors belonging to both the classes and the set of attributes is not empty, the process is repeated from step 1 on the new dataset with A eliminated from the set of features.

Output:
Decision tree classifier

Phase 1 (Determination of the completeness of lung): In this phase a decision tree classifier is used to determine if iterative thresholding and morphological operations are sufficient for complete segmentation of lungs:

Input:
Chest CT slice, I
Process logic:
Step 1: The image is converted to binary image by applying steps 1 through 5 of the dataset generation process.
Step 2: The feature vector is fed to the decision tree classifier built from the generated dataset.
Output:
Class label, 1 or 2 and the binary image, C.

If the output is 1, the lung segmented by iterative thresholding and morphological operations is incomplete and if the output is 2 iterative thresholding and morphological operations is complete.

Phase 2 (Extraction of lung parenchyma): In this phase, the lung parenchyma is extracted from the chest CT. The choice of operations in this phase depends on the output of phase 1. The major algorithm applied in this phase is ACM (Chan and Vese, 2001). The contribution lies in automated initialization of the contour for ACM, in that it combines the lung region obtained by supervised algorithm (Darmanayagam *et al*., 2013) and the convex hull concept used by the Convex Hull Based algorithm (Elizabeth *et al*., 2012a, b). It also introduces an additional constraint in the termination of ACM which ensures that the number of connected components does not exceed two.

Input:
Binary image, C and class label
Process logic:
Step 1: If class label is 2, go to step 8.
Step 2: The longer connected component, c_lung is chosen based on height of the individual regions in the binary image.
Step 3: The chosen connected component with height, h is folded horizontally about the centroid to obtain r_c_lung.

$$r\_c\_lung(x, y) = c\_lung(x, N\text{-}y) \quad (7)$$

where, N is the number of columns.

Step 4: The reflected image, r_c_lung is translated horizontally in such a way that the point of the corresponding region in the binary image C that is closest to the centre coincides with the point in the corresponding row of r_c_lung that is closest to the centre.

$$lung = (r\_c\_lung(x, y + dist))\,OR\,(c\_lung) \quad (8)$$

Where:

$$dist = \begin{cases} (\min(y) \text{ for which } l\_lung \text{ is } 1) - (\min(y) \text{ for which } r\_c\_lung \text{ is } 1), & \text{if right connected component is long} \\ [(\min(y) \text{ for which } r\_lung \text{ is } 1) + width(r\_lung)] - & \text{if left connected component is long} \\ [(\min(y) \text{ for which is } r\_c\_lung \text{ is } 1) + width(r\_c\_lung)], & \end{cases}$$

where, l_lung represents the left lung, r_lung represents the right lung as defined by C.

Step 5: The initial contour is extracted.
  a. Morphological open operation is applied on the image, lung, obtained in step 4.
  b. The black pixels not connected to the border are set to white.
  c. The centre point, (c1x, c1y) between the two lungs is determined using Eq. 9 and 10:

$$clx = \min(R) + \frac{\max(R) - \min(R)}{2} \quad (9)$$

$$cly = \min(C) + \frac{\max(C) - \min(C)}{2} \quad (10)$$

where, R is the list of coordinates along the vertical direction, i.e., the row index and C is the list of coordinates along the horizontal direction, i.e., the column index

  d. A rectangular mask of size:

$$h \times \frac{\max(C) - \min(C)}{2}$$

is constructed and placed over the image obtained in step 5 b, centered at (c1x, c1y); the region covered by the mask is filled with ones to get a single connected component.

e. The convex hull of the single connected component obtained in step 5 e is constructed.
f. The edge is removed from a region of size:

$$h \times \frac{\max(C) - \min(C)}{2}$$

centered at (c1x, c1y) and the convex outer edges alone are retained.
g. Linear interpolation is applied to the points in the convex hull to get a convex area without notches.
h. The edge of the convex area is obtained using canny filter.
i. The edge is removed from a region of size :

$$h \times \frac{\max(C) - \min(C)}{2}$$

centered at (c1x, c1y) and the convex outer edges alone are retained.
j. A logical OR operation is performed between the resultant edge and the binary image obtained in step 5 f.
k. The holes are filled.
l. Morphological erosion is applied.

Step 6: ACM algorithm is applied.
a. The eroded image is given as mask, C to the active contour model
b. The foreground and background pixels are identified
  i. The distance map of C, D is obtained
  ii. The distance map of (1-C), E is obtained
  iii. A matrix, ϕ is constructed using the distance map of the initial mask:

$$\phi(x,y) = D(x,y) - E(x,y) + C(x,y) - 0.5 \tag{11}$$

  iv. Initial force is set to eps, 2^{-52} (default in matlab), to avoid division by zero.
  v. The connected components of the mask are labeled.
  vi. The pixels with a negative value in ϕ are considered background pixels and the others are considered foreground pixels.
c. Compute the Heaviside step function of ϕ, H:

$$H(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1, & \text{for } x > 0 \\ 0.5, & \text{for } x = 0 \end{cases} \tag{12}$$

d. The average intensity of foreground pixels is computed:

$$C1 = \frac{\sum_x \sum_y I(x,y)H(x,y)}{N_f + \text{eps}} \tag{13}$$

e. The average intensity of background pixels is computed:

$$C2 = \frac{\sum_x \sum_y I(x,y)(1 - H(x,y))}{N_b + \text{eps}} \tag{14}$$

f. The energy is computed:

$$F(x,y) = -(I(x,y) - C1)^2 + (I(x,y) - C2)^2 \tag{15}$$

g. Curvature of ϕ, K is computed:

$$K(x,y) = \frac{\phi_{xx}(x,y)\phi_y^2(x,y) - 2\phi_{xy}(x,y)\phi_x(x,y)I_y(x,y) + \phi_{yy}(x,y)\phi_x^2(x,y)}{(\phi_x(x,y) + \phi_y(x,y))^{3/2}} \tag{16}$$

where, ϕ_x, ϕ_y and ϕ_{xy} are derivatives of ϕ obtained by forward finite difference:

$$\phi_x = \frac{\phi_{x+\Delta x, y} - \phi_{xy}}{\Delta x}$$

$$\phi_y = \frac{\phi_{x, y+\Delta y} - \phi_{x,y}}{\Delta y}$$

$$\phi_{xy} = \frac{(\phi_{x+\Delta x, y+\Delta y} - \phi_{x, y+\Delta y}) - (\phi_{x+\Delta x, y} - \phi_{x,y})}{\Delta x \Delta y}$$

with:

$\Delta x = \Delta y = 1$
$\phi_x = \phi_{x+1,y} - \phi_{x,y}$
$\phi_y = \phi_{x,y+1} - \phi_{x,y}$
$\phi_{xy} = (\phi_{x+1,y+1} - \phi_{x,y+1}) - (\phi_{x+1,y} - \phi_{x,y})$

h. The curvature is normalized.

$$\phi'(x,y) = \frac{\phi(x,y)}{\max(\phi(x,y))} \tag{17}$$

i. The external force of the image is calculated:

$$\text{Force}(x,y) = \mu\phi(x,y) + F(x,y) \tag{18}$$

where, μ is a positive length parameter.
j. The force is normalized by dividing by the maximum value of force.
k. The stepsize, h is set to 0.5 based on experimentation
1. ϕ is updated to ϕ_{new}.

$$\phi_{new}(x,y) = \phi(x,y) + h \times \text{Force}(x,y) \tag{19}$$

m. The average distance between ϕ and ϕ_{new} based on the points for which deviation is >0.5 is computed by:

$$Q = \sum_{|\phi_{new} - \phi| > 0.5} \frac{(\phi_{new} - \phi)}{M} \tag{20}$$

where, M is the No. of points for which deviation is >0.5.
n. The new mask, C is:

$$C(x,y) = \begin{cases} 1, & \phi(x,y) \le 0 \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

o. Ensure that the contour of the mask is stationary and the number of components does not exceed 2.
  i. if $Q \le 0.18^2 h$ or if number of connected components exceeds 2 go to step 7.
  ii. if $Q > 0.18^2 h$, repeat from step 6 b, with $\phi = \phi_{new}$.

Step 7: The locations in which C is 1 are considered to form the lung region.

Step 8: Lung parenchyma, L is extracted.

$$L(x,y) = \begin{cases} I(x,y), & \text{if } C(x,y) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

Output:
Lung parenchyma

EXPERIMENTAL RESULTS

A set of 70 chest CT image slices of different patients were taken for the experiment. The images were segmented using iterative thresholding and morphological operations. The relationship between the shape features of the left lung and that of the right lung can be used to determine the completeness of the segmented lung. In order to determine this relationship, three possible sets of shape features that could discriminate among the complete and incomplete lungs were identified. Three decision trees were built based on these three sets of features using CART algorithm and validated using leave one out method. Based on the features that were used by the three decision trees, an optimal set of features were determined and used for the construction of the fourth decision tree which has been used by the segmentation subsystem to determine the completeness of the lung in phase 1 of the Proposed Segmentation algorithm.

Decision tree 1: A feature vector consisting of the ratios of convexity, area, eccentricity, equiv diameter, height, width, perimeter, major axis length and minor axis length of left lung to the corresponding feature of the right lung is constructed for each of the 70 chest CT slices. The decision tree constructed using CART Method and tested using leave one out cross validation for the last fold is shown in Fig. 3. The rules derived are shown in Fig. 4. The error rate achieved is 11.43%.

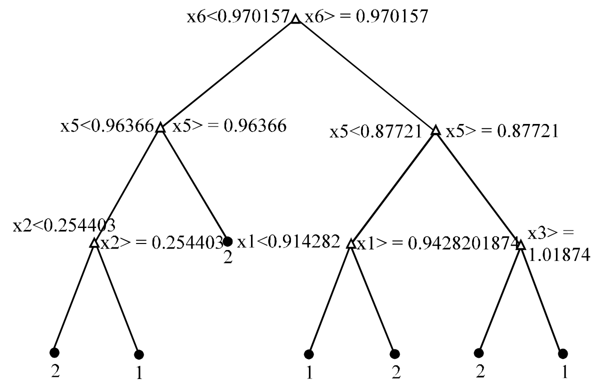

Fig. 3: Decision tree 1; x6: ratio of the width of left lung to that of right lung; x5: ratio of the height of left lung to that of right lung; x2: ratio of area of left lung to that of right lung; x1: ratio of convexity of left lung to that of right lung and x3: ratio of eccentricity of left lung to that of right lung

Classification rules based on decision tree 1:

1 if x6<0.970157 then node 2 elseif x6>=0.970157 then node 3 else 1
2 if x5<0.96366 then node 4 elseif x5>=0.96366 then node 5 else 1
3 if x5<0.87721 then node 6 elseif x5>=0.87721 then node 7 else 2
4 if x2<0.254403 then node 8 elseif x2>=0.254403 then node 9 else 1
5 class = 2
6 if x1<0.914282 then node 10 elseif x1>=0.914282 then node 11 else 2
7 if x3<1.01874 then node 12 elseif x3>=1.01874 then node 13 else 1
8 class = 2
9 class = 1
10 class = 1
11 class = 2
12 class = 2
13 class = 1

Decision tree 2: A feature vector consisting of convexity of left lung-convexity of right lung, area of left lung-area of right lung, eccentricity of left lung-eccentricity of right lung, equiv diameter of left lung-equiv diameter of right lung, height of left lung-height of right lung, width of left lung-width of right lung, perimeter of left lung-perimeter of right lung, major axis length of left lung-major axis length of right lung and minor axis length of left lung-minor axis length of the right lung is constructed for each of the 70 chest CT slices. The decision tree constructed using CART method and tested using leave one out cross validation for the last fold is shown in Fig. 4. The rules derived are discuss in study. The error rate achieved is 21.43%.

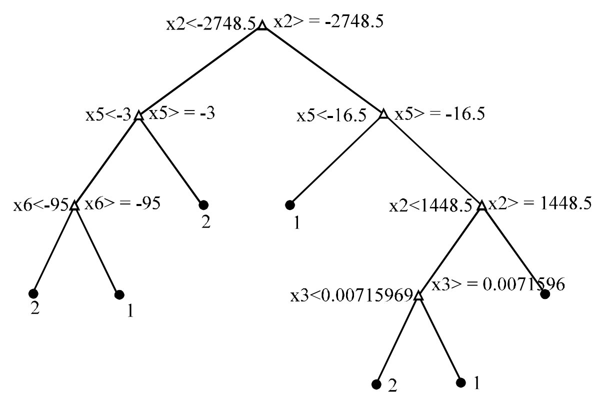

Fig. 4: Decision tree 2; x2: area of left lung-area of right lung; x5: height of left lung-height of right lung; x6: width of left lung-width of right lung and x3: eccentricity of left lung-eccentricity of right lung

Classification rules based on decision tree 2:

1. if x2<-2748.5 then node 2 elseif x2>=-2748.5 then node 3 else 1
2. if x5<-3 then node 4 elseif x5>=-3 then node 5 else 1
3. if x5<-16.5 then node 6 elseif x5>=-16.5 then node 7 else 2
4. if x6<-95 then node 8 elseif x6>=-95 then node 9 else 1
5. class = 2
6. class = 2
7. if x2<1448.5 then node 10 elseif x2>= 448.5 then node 11 else 1
8. class = 2
9. class = 1
10. if x3<0.00715969 then node 12 elseif x3>=0.00715969 then node 13 else 1
11. class = 2
12. class = 2
13. class = 1

Decision tree 3: A feature vector consisting of convexity of the individual lungs, ratio of convexity of the left lung to that of the right lung, area of the individual lungs, ratio of the area of the left lung to that of the right lung, eccentricity of the individual lungs, ratio of the eccentricity of the left lung to that of the right lung, equiv diameter of the individual lungs, ratio of the equiv diameter of the left lung to that of the right lung, height of the individual lungs, ratio of the height of the left lung to that of the right lung, width of the individual lungs, ratio of the width of the left lung to that of the right lung, perimeter of the individual lungs, ratio of the perimeter of the left lung to that of the right lung, major axis length of the individual lungs, ratio of the major axis length of the left lung to that of the right lung, minor axis length of the individual lungs and ratio of the minor axis length of the left lung to that of the right lung is constructed for each of the 70 chest CT slices. The decision tree constructed using CART Method and tested using leave one out cross validation for the last fold is shown in Fig. 5. The rules derived are shown in study. The error rate achieved is 10%.

Classification rules based on decision tree 3:

1 if x1<0.807568 then node 2 elseif x1>=0.807568 then node 3 else 1
2 if x17<0.596114 then node 4 elseif x17>=0.596114 then node 5 else 2
3 if x4<4319 then node
4 class = 1
5 class = 2
6 class = 2
7 class = 1

Decision tree 4: Based on the observations from decision trees 1, 2 and 3, a feature vector consisting of the optimal set of features, namely, ratios of width, height, area, convexity and eccentricity of the left lung region to that of the right lung region and the convexity, width, height, area and eccentricity of the individual lungs is constructed for each of the 70 chest CT slices. The decision tree constructed using CART Method and tested using leave one out cross validation for the last fold is shown in Fig. 6. The rules derived are discuss in the study. The error rate achieved is 8.57%.

Classification rules based on decision tree 4:

1 if x7<0.807568 then node 2 elseif x7>=0.807568 then node 3 else 1
2 if x2<0.596114 then node 4 elseif x2>=0.596114 then node 5 else 2
3 if x1<1.32699 then node 6 elseif x1>=1.32699 then node 7 else 1
4 class = 1
5 class = 2
6 class = 1
7 class = 2

The error rate achieved by the decision tree constructed using the optimal set of features is less than that achieved by the other three decision tree classifiers. Hence, decision tree 4 has been used by phase 1 of the proposed Two-Phase Segmentation algorithm for segmentation of lungs in chest CT.

The proposed Two Phase Segmentation algorithm has been tested with chest CT images of subjects affected by various types of lung disorders, namely, bronchiectasis, tuberculosis and pneumonia. It has been compared with the conventional Iterative Thresholding Method, Convex Hull Based algorithm and Supervised algorithm for segmentation. Figure 7-9 show the results obtained for chest CTs of patients affected by bronchiectasis, tuberculosis and pneumonia, respectively.

Quantitative evaluation of the accuracy of the Lung Segmentation algorithm has been done by computing the overlap percentage. The overlap percentage is defined by Eq. 23:

$$\text{Overlap} = \left|\frac{N_{seg} \cap N_{man}}{N_{seg} \cup N_{man}}\right| \tag{23}$$

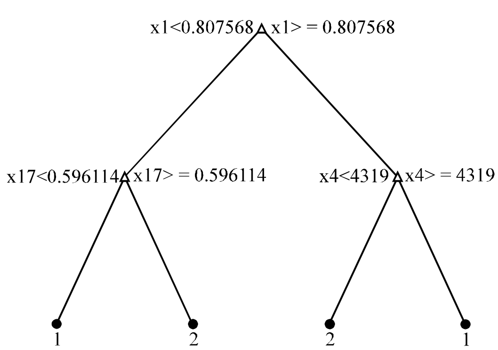

Fig. 5: Decision tree; 3x1: convexity of right lung; x17: area of right lung and x4: ratio of the height of left lung to that of right lung

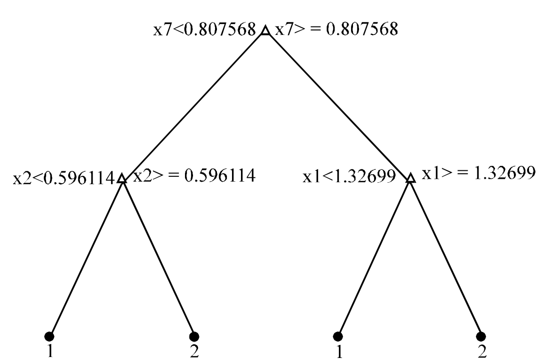

Fig. 6: Decision tree 4; x7: convexity of the right lung; x2: width of 1 eft lung/height of right lung and x1: height of left lung/height of right lung

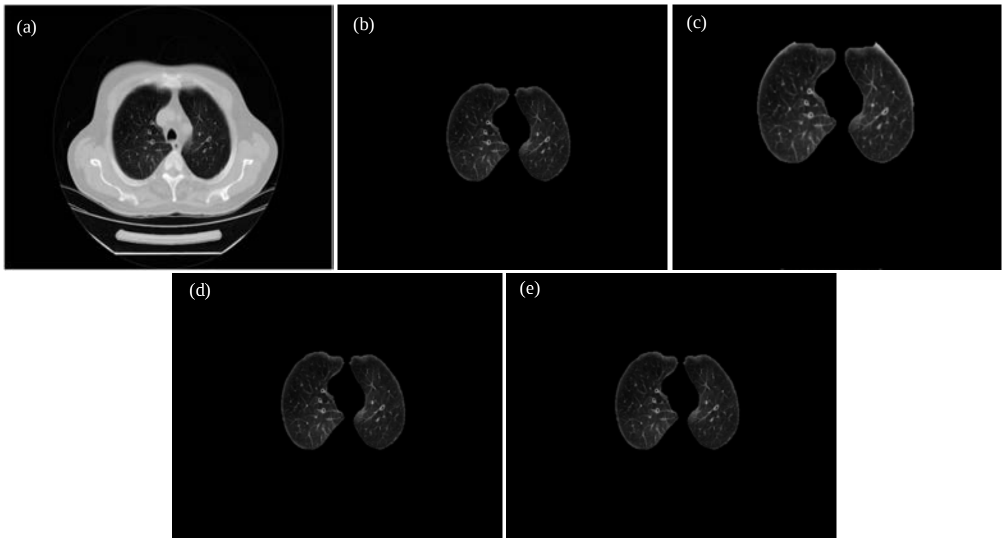

Fig. 7: Results obtained for chest CT of a patient affected by bronchiectais; a) Chest CT; b) Thresholding Based algorithm; c) Convex Hull Based algorithm; d) Supervised algorithm and e) Proposed Two Phase Supervised algorithm

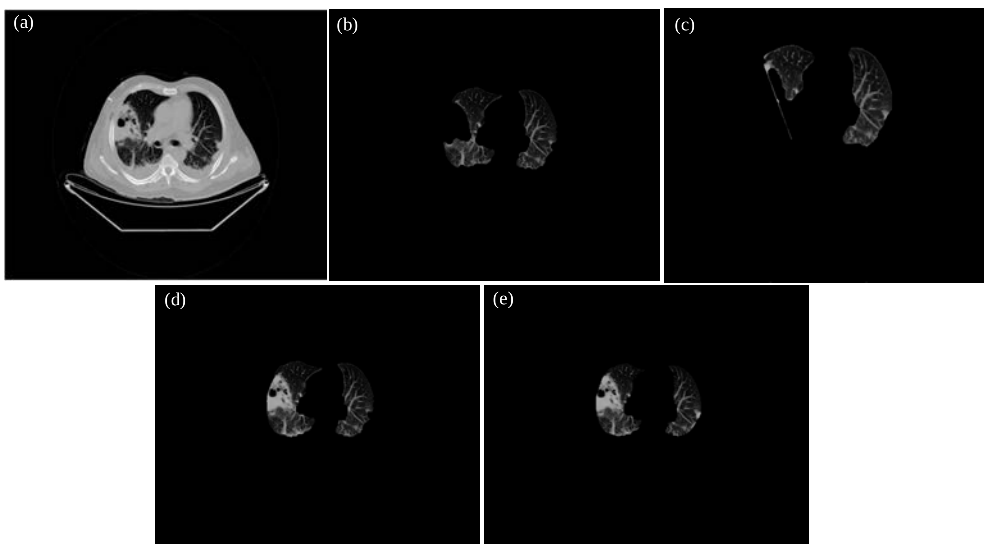

Fig. 8: Results obtained for chest CT of a patient affected by tuberculosis; a) Chest CT; b) Thresholding Based algorithm; c) Convex Hull Based algorithm; d) Supervised algorithm and e) Proposed Two Phase Supervised algorithm

where, $N_{seg} \cap N_{man}$ denotes the number of pixels in common between the lung segmented by the Segmentation algorithm and the manually segmented lung and $N_{seg} \cap N_{man}$ denotes the number of pixels in that belong to the lung region in the lung segmented by the segmentation algorithm and the manually segmented lung. The best case is 100 which means in cases where there are no juxta-pleural nodules, the discontinuity between the lung and the outer chest region is clear and hence the lung segmented by all the algorithms considered in this work perform effectively with 100% overlap. The worst case is a overlap of 55.3% with the proposed Two Phase Supervised Lung Segmentation algorithm, 37.83% with iterative thresholding based segmentation, 25.82% with convex-hull based segmentation and 54.25% with supervised segmentation. Thus, the proposed algorithm is found to perform better as compared to the other three algorithms with which it has been compared.

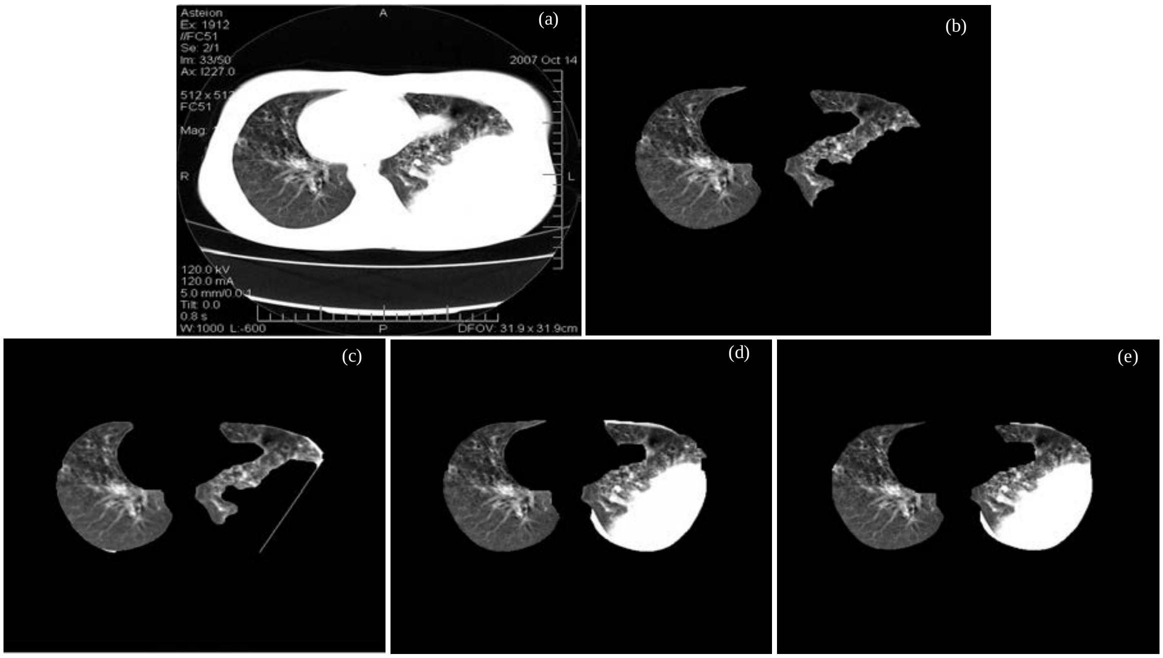

Fig. 9: Results obtained for chest CT of a patient affected by pneumonia; a) chest CT; b) Thresholding Based algorithm; c) Convex Hull Based algorithm; d) Supervised algorithm and e) Proposed Two Phase Supervised algorithm

CONCLUSION

In this study, a novel supervised Two Phase Segmentation algorithm has been proposed for segmentation of lungs from chest CT images. This has been found to be a challenging task especially in cases of subjects whose lung consists of peripherally placed PBRs as discussed in literature (Wei *et al*., 2013; Darmanayagam *et al*., 2013; Elizabeth *et al*., 2012b). The proposed algorithm combines the advantages of Iterative Thresholding Method, Convex Hull Based algorithm, Supervised algorithm and ACM that have been used in literature for segmentation of lungs and hence found to produce better segmentation results as compared with each of the individual algorithms.

The Two Phase Supervised Segmentation algorithm has been developed to suit the segmentation step in computer aided diagnosis of lung disorders. It has been tested with chest CTs of subjects infected by bronchiectasis, tuberculosis and pneumonia and has been found to produce a minimum overlap of 55.3% against 37.83% achieved with iterative thresholding based segmentation, 25.82% achieved with convex-hull based segmentation and 54.25% achieved with supervised segmentation. Segmentation of lungs is a crucial and challenging step in CAD of lung disorders. Hence, eventhough better results have been achieved; further research is required to produce accurate segmentation results.

REFERENCES

Armato, S.G. and W.F. Sensakovic, 2004. Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis. Acad. Radiol., 11: 1011-1021.

Breiman, L., J. Friedman, R.A. Olshen and C.J. Stone, 1984. Classification and Regression Trees. Wadsworth International Group, USA., ISBN-10: 0534980546, Pages: 368.

Chan, T.F. and L.A. Vese, 2001. Active contours without edges. IEEE Trans. Image Process., 10: 266-277.

Collins, V.P., R.K. Loeffler and H. Tivey, 1956. Observations on growth rates of human tumors. Am. J. Roentgenol. Radium Ther. Nuclear Med., 76: 988-1000.

Darmanayagam, S.E., K.N. Harichandran, S.R.R. Cyril and K. Arputharaj, 2013. A novel supervised approach for segmentation of lung parenchyma from chest CT for computer-aided diagnosis. J. Digit. Imag., 26: 496-509.

Elizabeth, D.S., A. Kannan and H.K. Nehemiah, 2009. Computer-aided diagnosis system for the detection of bronchiectasis in chest computed tomography images. Int. J. Imaging Syst. Technol., 19: 290-298.

Elizabeth, D.S., H.K. Nehemiah, C. Raj and A. Kannan, 2012a. A novel segmentation approach for improving diagnostic accuracy of CAD systems for detecting lung cancer from chest computed tomography images. J. Data Inform. Qual., 3: 1-4.

Elizabeth, D.S., H.K. Nehemiah, R. Raj and A. Kannan, 2012b. Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image. IET Image Process., 6: 697-705.

Nathan, M.H., V.P. Collins and R.A. Adams, 1962. Differentiation of benign and malignant pulmonary nodules by growth rate. Radiology, 79: 221-232.

Otsu, N., 1979. A threshold selection method from gray-level histogram. IEEE Trans. Syst. Man Cybern., 9: 62-66.

Ridler, T. and S. Calvard, 1978. Picture thresholding using an iterative selection method. IEEE Trans. Syst. Man Cybernetics, 8: 630-632.

Wei, Y., G. Shen and J.J. Li, 2013. A fully automatic method for lung parenchyma segmentation and repairing. J. Digital Imaging, 26: 483-495.