Automatic Vessel Segmentation of Lung Affected Patterns in MDCT Using Decision Tree Classification

Jithya S. babu and Sreeja Mole

Jithya S. babu Department of ECE, Sivaji College of Engineering and Technology,
Manivila, Kanyakumari Dist, Tamilnadu, India

Abstract: Pulmonary vascular tree segmentation is gaining importance since it is one of the fundamental basis for different applications, such as the detection of interstitial pneumonia (IP), pulmonary emboli etc. Such an application will require an accurate and reliable segmentation of pulmonary vessels. The accuracy of this preprocessing stage is bound to influence the accuracy in computer aided diagnosis (CAD) of IP patterns. While lot many algorithms aimed at improving accuracy of lung segmentation, vessel tree segmentation is still an open research issue. In this paper an automated vessel tree segmentation algorithm with high accuracy is proposed in presence of pathologies affecting lung parenchyma. The initial stage accounts for a vessel enhancement filtering, which uses second order local structure of an image (Hessian) with vessel ness measure obtained on the basis of all eigen values of the Hessian. Followingly texture based refinement using 3D co-occurrence matrix, which ties all textures together (a single value per feature will take into account all texture features within a lung, finally the classification is performed to correct possible over segmentation. The proposed method utilizes Decision Tree classifier which improves traditional SVM by adding membership to training sample to indicate degree of membership of this sample to different class. Consequently it reduces noises and outliers in data and enhances performance and accuracy of SVM. The performance of the proposed scheme and of the previously reported technique, in vessel tree segmentation was evaluated by means of area overlap; true positive fraction and false positive fraction of image dataset obtained from IP affected patient scans. The method is expected to improve the performance as compared to other reported techniques.

Key words: Pulmonary vascular tree segmentation · Interstitial pneumonia and computer aided diagnosis · Decision Tree · Support Vector Machine classifier · Image enhancement, hessian matrix

INTRODUCTION

Interstitial lung disease (ILD) refers to a group of diseases affecting the interstitial (tissue and space around the air sacs of the lungs). Usual interstitial pneumonia (UIP) is a form of lung disease characterized by progressive scarring of both lungs shown in Fig. 1. UIP may be diagnosed by a radiologist using a computed tomogram of the chest, or by a pathologist using tissue obtained by a lung biopsy. Radiologic ally, the main feature required for a confident diagnosis of UIP is honeycomb change in the periphery and the lower portions (bases) of the lungs. For lung imaging and analyzing, computed tomography (CT) is widely used;

In contrast to high resolution CT scanning, which allows only a limited portion of lung parenchyma to be sampled, MDCT (Multi Detector CT) which permits CT scanners to acquire multiple slices or sections simultaneously and greatly increases speed, allowing acquisition visualization, characterization and quantification of the entire extend of lung anatomy, which aids in the analysis of lung malformations.

Recently MDCT analysis of ILD affected has been introduced [1]. These methods [2-5] mainly focused based on the texture analysis for identification and characterization.There are different algorithms already developed for improving the accuracy of lung fields in presence of ILD. But corresponding vessel tree

Corresponding Author: Jithya S. Babu, Department of ECE, Sivaji College of Engineering and Technology,
Manivila, Kanyakumari Dist, Tamilnadu, India.

Fig. 1: Figure depicting honey combing in patient with Interstitial Pneumonia

segmentation is still an open research issue, due to the complexity of vessel tree, diversity of vessel size and intensity, presence of noise etc. The previously proposed methods deals with radiologic appearance of normal lung parenchyma

[10-16], with focal abnormalities [17-22] and with pulmonary embolism [23].In another case of lung image registration [24], airway tree [11] and lung lobe segmentation [25], similar vessel tree segmentation is Applied.

In this paper an automated vessel tree segmentation scheme with improved accuracy rate is proposed to deal with IP affected lung parenchyma. The method is applied to volumetric scans of patients affected by interstitial pneumonia. The algorithm deals with a 3D multiscale vessel enhancement filtering based on eigen value analysis of hessian matrix and on supervised segmentation.

Literature Review: MDCT based identification of IP is reported in [1]. The algorithms developed for the accuracy of lung field segmentation is reported in [7-9]. The vessel tree segmentation method which were reported in [10-16] deals with appearance of normal lung parenchyma, along with focal abnormalities in [17-22] and with pulmonary embolism [23]. But the problem with this method is that they mainly rely on single or multiscale image enhancement combined with threshold reported in literatures [11-20], [22] or with of volumetric data sets with almost isotropic voxels, enabling unsupervised segmentation [23] to enhance the tubular vascular structures. Followingly other methods such as region growing [7], level sets [14] and fuzzy connectedness havebeen reported. Their drawback is their parametric nature. In case of preprocessing steps in lung image registration [2], airwaytree [11] and lung lobe

segmentation [25], similar vessel tree segmentation method is used. Kollar *et al* [17] introduced single scale enhancement filter and adopted in [4] without however capturing varying size of vessel tree segments. Multiscale approaches which mainly based on eigen value analysis of Hessian matrix are exploited by Frangi *et al* [20], Sato *et al* [18-22], Li *et al* [19], Agam *et al* [2], Krissan *et al* [12], Zohu *et al* [23] and Lo *et al* [13] employing different response filter. To distinguish between vessel tree and noise components Shukta *et al* [1] proposed multiscale technique with connected component analysis and branch point analysis.

MATERIALS AND METHODS

The proposed method consist of five modules which are.

- Image acquisition
- Wavelet Edge Enhancement
- Lung segmentation
- Vessel tree volume Identification
- Feature Extraction
- Decision Tree classification

The flow diagram of the method is depicted in Fig 2.

Image Acquisition: This module is contributed to acquire more number of samples from patients detected with IP patterns and from normal patients using MDCT scanner.

Lung Segmentation: The proposed methods exploits advantages offered by a 2D wavelet pre-processing step and core of method is automated 3D histogram thresholding [11]. Thresholding combined with Wavelet Edge Enhancement is successfully used in lung field
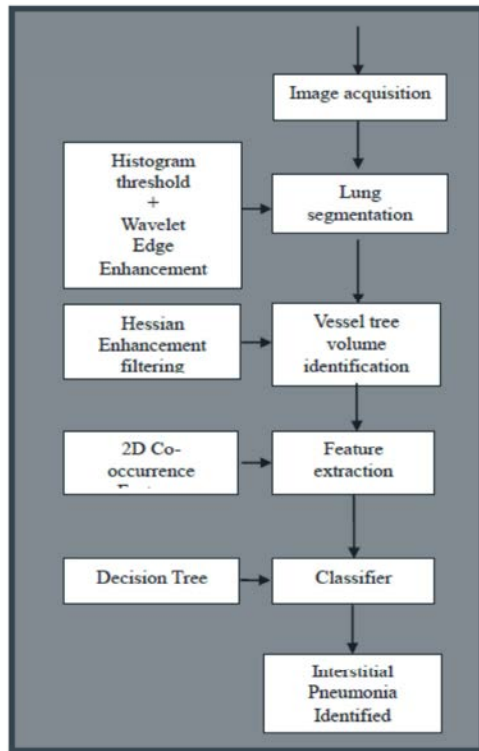
Fig. 2: Flow diagram of proposed method

segmentation by Korfiatis [12]. Strong edges generate large wavelet coefficients at all levels, which is not true for weak edges. As a result multiplying the coefficients across the scales enhances the large coefficients more than the smaller coefficients in the correlation image. To solve this problem we compress the dynamic range of the correlation image. However, gray level-based algorithms are insufficient in correctly segmenting lung fields in case of IPs affecting lung borders, since IPs are manifested as tissue texture alterations. To overcome this LF under-segmentation, a texture based border refinement step is employed mentioned by [9, 10].

Vessel Tree Volume Identification: To increase the effectiveness of vessel segmentation algorithm vessel enhancement procedures are first applied as a preprocessing step [17-23]. Here we are using a hessian- based vessel enhancement method which uses eigen value of hessian matrix to distinguish vessels from background exploiting a 3D-tubular structure associated to vessel tree.

To meet with the wide range of vessel sizes, original images are convolved with Gaussian kernels of varying standard deviation enhancing local structures of specific sizes, followed by combination of the local maxima of

filter responses at multiple scales. Calculating the eigenvalues [l1-l3] of the Hessian matrix of each voxel, the response of the tubular structures is approximated according to Zhou *et al.* [23] By:

$$R(x,y,z,\sigma_s;l_1,l_2,l_3) =$$

$$\left\{ \begin{array}{ll} \dfrac{(l_1+l_2)}{2} * \exp\left(-\left|\dfrac{|l_1|}{\sqrt{l_1^2+l_2^2+l_3^2}}-c\right|\right), & l_1,l_2,l_3 < 0 \\ 0 & \text{otherwise} \end{array} \right\}$$

A value of 0.7 is adopted for parameter c by [23]. The filter responses at each scale are normalized to achieve a fair comparison among multiple scales. Considering vessel tree size varying from 2 to 24 mm, Gaussian kernels with standard deviation ranging from 1 to 12 voxels were utilized.An Expectation Maximization (EM) segmentation algorithm is then applied to the filter response volumes in order to identify the voxels with high responses associated with tubular structures [23]. Followingly we apply the EM segmentation algorithm at all scales and a hierarchical scheme is implemented to combine the results across the scales, providing vessel tree volume candidate.

Feature Extraction: The refinement of the vessel tree is obtained by a classifier based on 3D texture analysis, which uses 3D co-occurrence features. 3Dco-occurrence matrices are matrices that are able to capture the spatial dependence of gray-level values across multiple slices, whereas the two- dimensional co-occurrence matrices capture the spatial dependence of gray levels within a specific slice (scan). Gray level co-occurrence matrix (GLCM) [26] is a well-established tool for characterizing the spatial distribution (second order statistics) of gray levels in an image and has been extensively exploited in lung image analysis [27]. GLCMs were generated for 13 directions and two distances (d = 1, 2 pixels). Thirteen second order statistics (angular second moment, contrast correlation, variance, inverse different moment, sum average, sum, variance, sum entropy, entropy, difference variance, difference, entropy, information measure of correlation 1 and information measure of correlation 2) were extracted from each GLCM. The mean and range values of each second order statistic over the 13 directions were calculated resulting in a total of 52 features.

Disciminant analysis which is used in statistics, pattern recognition etc is preferred for feature extraction process and dimensionality reduction of initial 52 features.

The goal of SDA is to sequentially identify those variables (features) that widely separate the classes from one another while keeping the classes themselves as tightly clustered as possible. A feature set of four features was selected consisting of: Mean of Variance (d =1pixels), Range of Sum Average (d = 1pixels), Mean of Sum Entropy (d = 2 pixels) and Mean of Variance (d = 2 pixels).

Decision Tree Classification: A decision tree is a decision support tool that uses a tree- like graph or model of decisions and their possible consequences, including chanceevent outcomes, resource costs and utility. It is oneway to display an algorithm.Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. An example is shown on the right. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions. This process oftop-down induction of decision trees (TDIDT) [1] is an example of a greedy algorithm and it is by far the most common strategy for learning decision trees from data, but it is not the only strategy. In fact, some approaches have been developed recently allowing tree induction to be performed in a bottom-up fashion. [2] In data mining, decision trees can be described also as the combination of mathematical and computational techniques to aid the description, categorisation and generalisation of a given set of data.

Data comes in records of the form:

$$(x, Y) = (x_1, x_2, x_3,....,x_k, Y)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalise. The vector x is composed of the input variables, x1, x2, x3 etc., that are used for that task.

Decision trees used in data mining are of two main types:

- Classification tree analysis is when the predicted outcome is the class to which the data belongs.
- Regression tree analysis is when the predicted outcome can be considered a real number (e.g. the price of a house, or a patient's length of stay in a hospital).

The term Classification and Regression Tree (CART) analysis is an umbrella term used to refer to both of the above procedures, first introduced by Breiman *et al.* [3] Trees used for regression and trees used for classification have some similarities-but also some differences, such as the procedure used to determine where to split [3].

Some techniques, often called ensemble methods, construct more than one decision tree:

- Bagging decision trees, an early ensemble method, builds multiple decision trees by repeatedly resampling training data with replacement and voting the trees for a consensus prediction [4].
- A Random Forest classifier uses a number of decision trees, in order to improve the classification rate.
- Boosted Trees can be used for regression-type and classification-type problems [5, 6].
- Rotation forest-in which every decision tree is trained by first applying principal component analysis (PCA) on a random subset of the input features [7].

Decision tree is the learning of decision tree from class labeled training tuples. A decision tree is a flow chart like structure, where each internal (non-leaf) node denotes a test on an attribute, each branch represents an outcome of the test and each leaf (or terminal) node holds a class label. The topmost node in tree is the root node.

There are many specific decision-tree algorithms. Notable ones include:

- ID3 (Iterative Dichotomiser 3)
- C4.5 algorithm, successor of ID3
- CART (Classification And Regression Tree)
- CHi-squared Automatic Interaction Detector (CHAID). Performs multi-level splits when computing classification trees.[8]
- MARS: extends decision trees to better handle numerical data

C4.5 Algorithm: C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification and for this reason, C4.5 is often referred to as a statistical classifier. C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set consists of already classified samples. Eachsample ofap-dimensional vector, where therepresent attributes or features of the sample, as well as the class in whichfalls.At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterionis the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sublists.

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

Pseudocode: In pseudocode, the general algorithm for building decision trees is: [2]

- Check for base cases
- For each attribute a
- Find the normalized information gain from splitting on a
- Let a_best be the attribute with the highest normalized information gain
- Create a decision node that splits on a_best
- Recurse on the sublists obtained by splitting on a_best and add those nodes as children of node

Performance Evaluation: Segmentation accuracy of the proposed method was evaluated by means of area overlap (AO), true positive fraction (TPF) and false positive fraction (FPF) metrics. Due to large volume of data analyzed, the definition of voxel-exact ground truth of the vessel tree volume, required for quantitative
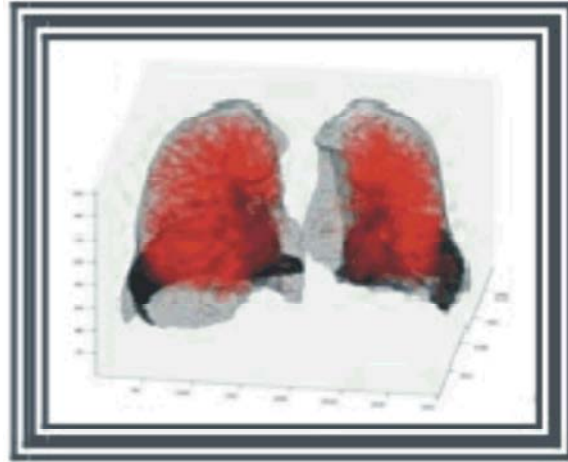

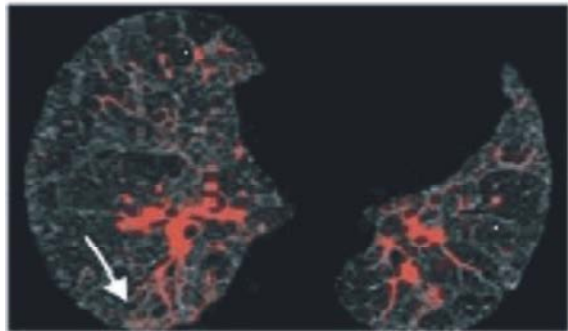Fig. 3: Vessel tree segmentation 3-D representation

Fig. 4:

evaluation of the algorithm segmentation accuracy, is a tedious task.In Shikata *et al*. [11] and Zhou *et al*. [23] evaluation was performed by means of control points tracking the center lines of vessels, provided by two radiologists using a Graphical User Interface (GUI). Both studies recognized the difficulties in creating a pixel-exact ground truth attributed to the fuzziness of vessel tree segments due to partial volume effect and noise. In present ground truth can be derived by means of a GUI designed to facilitate editing of 2-D vessel segments. The GUI allowed the radiologist to review the original data in coronal, sagittal and axial planes and draw vessel tree segments.

DISCUSSION

In this study, an automated vessel tree segmentation scheme is reported dealing with IP affected lung parenchyma, as depicted in MDCT shown in Fig. 3, 4, 5. The development of vessel tree segmentation algorithms in case of IP-affected lung parenchyma is an open issue, challenged by the radiologic similarity of reticular
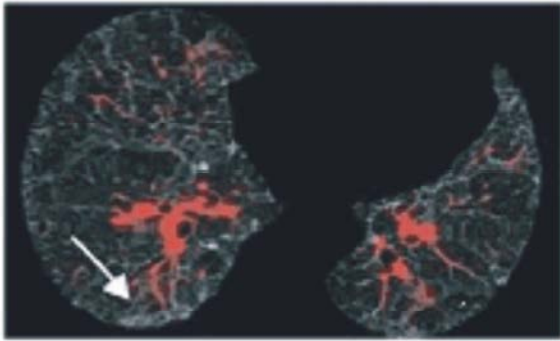
Fig. 5: Vessel tree segmentation- axial slice of segmented vessels after refinement

Fig. 6: Wavelet Edge Highlighting

Fig. 7: Vessel tree segmentation

patterns to vessel tree segments. While the majority of the proposed filters were limited to specific structures, this filter response has been designed in a way that enhances vessel tree segments and vessel bifurcations and in the same time suppresses non vessel structures. Furthe rmore, Zhou *et al*. [23] applied their technique on non contrast patient scans, similar to the ones exploited in this study, reporting high performance. Segmented vessel tree is depicted by red overlay. Arrows indicate vessel tree

over-segmentation example. To the best of the knowledge, this is the first vessel tree segmentation algorithm that is adapted to reticular patterns affecting lung parenchyma. This adaptation is attributed to the supervised fuzzy classification mechanism [31] incorporated in the second stage of the proposed method. The quantitative metrics area overlap (AO), true positive fraction (TPF) and false positive fraction (FPF) were considered by comparing the area of computer-derived borders to the ones derived by an expert radiologist [5]. For each accuracy segmentation metric, the mean (Mean), standard deviation (SD), minimum (Min), 1st quartile (Q1), Median, 3rd quartile (Q3) and maximum (Max) values were calculated.

ACKNOWLEDGMENT

REFERENCE

1. Aziz, Z.A., A.U. Wells, D.M. Hansell, S.J. Copley, S.R. Desai, S.M. Ellis, F.V. Gleeson, S. Grubnic, A.G. Nicholson, S.P.G. Padley, K.S. Pointon, J.H. Reynolds, R.J.H. Robertson and M.B. Rubens, 2004. RCTdiagnosis of diffuse parenchymal lung disease: Inter-observer variation,? Thorax, 59(6): 506-511.

2. Sluimer, I.C., P.F. van Waes, M.A. Viergever and B. van Ginneken, 2003. Computer-aided diagnosis in high resolution CT of the lungs,? Med. Phys., 30(12): 3081-3090.

3. Xu, Y., E.J.R. van Beek, Y. Hwanjo, J. Guo, G. McLennan and E.Hoffman, 2006. Computer-aided classification of interstitial lung diseasesvia MDCT: 3D adaptive multiple feature method (3D AMFM),? Acad. Radiol., 13(8): 969-978.

4. Zavaletta, V.A., B.J. Bartholmai and R.A. Robb, 2007. High resolutionmultidetector CT-aided tissue analysis and quantification of lung fibrosis,? Acad. Radiol., 14(7): 772-787.

5. Korfiatis, P., A. Karahaliou, A. Kazantzi, C. Kalogeropoulou and L. Costaridou, 2010. Towards quantification of interstitial pneumonia patterns in lung multidetector CT,? IEEE Trans. Inf. Technol. Biomed., 14(7): 675-680.

6. Armato S.G. and W.F. Sensakovic, 2004. Automated lung segmentation for thoracic CT Impact on computer-aided diagnosis,? Acad. Radiol., 11(9): 1011-1021.

7. Marten, K., V. Dicken, C. Kneitz, M. Hoehmann, W. Kenn, D. Hahn and C. Engelke, 2009. Computer-assisted quantification of interstitial lung disease associated with rheumatoid arthritis: Preliminary technical validation,?Eur. J. Radiol., 72(2): 278-83.

8. Korfiatis, P., S. Skiadopoulos, P. Sakellaropoulos, C. Kalogeropoulou and L. Costaridou, 2007. Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT,? Br. J. Radiol., 80(96): 996-1004.

9. Korfiatis, P., C. Kalogeropoulou, A. Karahaliou, A. Kazantzi, S. Skiadopoulos and L. Costaridou, 2008. Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT,? Med.Phys., 35(12): 5290-5302.

10. Sluimer, I.C., M. Prokop and B. van Ginneken, 2005. Towards automated segmentation of the pathological lung in CT,? IEEE Trans. Med. Imag., 24(98): 1025-1038.

11. Shikata, H.G. and E.A. McLennan, 2009. Hoffman, M. Sonka, ¯Segmentation of pulmonary vascular trees from thoracic 3-d ct images,? J. Biomed.Imag., 636240.

12. Krissian, K., G. Malandain, N. Ayache, R. Vaillant and Y. Trousse, 2000. Model-based detection of tubular structures in 3-d images,? Comput. Vis. Image Und., 80(2): 130-171.

13. Lo, P., B. van Ginneken and M. de Bruijne, 2010. Vessel tree extraction using locally optimal paths,? in Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro, pp: 680-683.

14. Vasilevskiy, A. and K. Siddiqi, 2001. Flux maximizing geometric flows,? IEEE Trans. Pattern Anal. Mach. Intell., 24(12): 1565-1578.

15. Kaftan, N., A.P. Kiraly, A. Bakai, M. Das, C.L. Novak and T. Aach, 2008. Fuzzy pulmonary vessel segmentation in contrast enhanced ct data,? in Medical maging: Image Processing, J. M. Reinhardt and J.P.W.

16. Wan, S.Y.L. Ritman and W.E. Higgins, 2008 Multi-generational analysis and visualization of the vascular tree in 3-D micro-CT images,? Comput.Biol. Med., 32(2): 55-71.

17. Koller, T., G. Gerig, G. Szekely and D. Dettwiler, 1995. Multiscale detection of curvilinear structures in 2-D and 3-D image data,? in Proc. Int. Conf.Comput. Vis., pp: 864-869.

18. Sato, Y., C. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura and R. Kikinis, 2000. Tissue classication based on 3-D local intensity structure for volume rendering,? IEEE Trans. Vis. Comput. Graph., 6(2): 160-180.

19. Li, Q., S. Sone and K. Doi, 2003. Selective enhancement filters for nodules vessels and airway walls in two- and three-dimensional CT scans,? Med. Phys., 30(8): 2040-2051.

20. Frangi, A., W. Niessen, K. Vincken and M. Viergever, 1998. Multiscale vessel enhancement filtering,? Lect. Notes Comput. Sc., 14(96): 130-137.

21. Agam, G., S.G. Armato, III and C. Wu, 2005. Vessel tree reconstruction in thoracic ct scans with application to nodule detection,? IEEE Trans. Med. Image., 24(4): 486-499.

22. Sato, Y., S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig and R. Kikinis, 1998. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images,?Med. Image Anal., 2(2): 143-168.

23. Zhou, C., H.P. Chan, B. Shahiner, L.M. Hadjiiski, A. Chughtai, S. Patel, J. Wei, J. Ge, P.N. Cascade and E.A. Kazerooni, 2007. Automatic multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications,? Med. Phys, 34(12): 4567-4577.

24. Yin, Y., E.A. Hoffman and C.L. Lin, 2009. Mass preserving nonrigid registration of CT lung images using cubic B-spline,? Med. Phys., 36: 4213-4222.

25. Ukil, S. and J.M. Reinhardt, 2009. Anatomy-guided lung lobe segmentation in X-ray CT images,? IEEE Trans. Med. Imag., 28(2): 02-214.

26. Haralick, R.M., K. Shanmugam and I.H. Dinstein, 1973. Textural features for image classification,? IEEE Trans. Syst. Man Cybern., vol. SMC-3, 6: 610-621.

27. Mariolis, I., P. Korfiatis, C. Kalogeropoulou, D. Daoussis, T. Petsas and L. Costaridou, 2010. Computer aided diagnosis of diffuse lung disease in multidetector CT—Selecting 3D texture features,? IFMBE Proc. 1,29(2): 208-211.

28. Einslein, K., A. Ralston and H.S. Wilf, 1977. Eds., Statistical Methods for Digital Computers. New York: Wiley, pp: 76-94.

29. Vapnik, N.V., 1995. The Nature of Statistical Learning Theory. NewYork:Springer-Verlag, [30] Chen, S., S. Zhou, F.-F. Yin, L. B. Marks and S. K. Das, 2007.Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis,? Med. Phys., 34(10): 3808-3814.

31. Chun-Fu Lin and Sheng-De Wang, 2002. Fuzzy support vector machines? IEEE Trans. Neural Netw. 13: 2.