Towards Quantification of Interstitial Pneumonia Patterns in Lung Multidetector CT

Panayiotis Korfiatis, Anna Karahaliou, Alexandra Kazantzi, Cristina Kalogeropoulou and Lena Costaridou

***Abstract*—Quantification of Diffuse Parenchyma Lung Disease (DPLD) patterns challenges Computer Aided Diagnosis schemes in Computed Tomography (CT) lung analysis. In this study, an automated scheme for volumetric quantification of Interstitial Pneumonia (IP) patterns, a subset of DPLDs, is presented, utilizing a MultiDetector CT (MDCT) data set. Initially, Lung Field (LF) segmentation is achieved by 3D automated gray level thresholding combined to wavelet highlighting, followed by a texture based border refinement step. The vessel tree volume is identified and removed from LF, resulting in Lung Parenchyma (LP) volume. Following, the abnormal LP is differentiated from normal LP utilizing a 2 class *k*-means clustering. Quantification of IP patterns is formulated as a three-class pattern recognition problem to classify abnormal LP into ground glass, reticular and honeycomb patterns, by means of SVM voxel classification, exploiting 3D co-occurrence features. Performance of the proposed scheme in segmenting LF, as well as in quantifying normal LP, ground glass, reticular and honeycomb patterns was evaluated by means of volume overlap on 5 MDCT scans. Volume overlap for left LF and right LF was 0.95±0.03 and 0.96±0.02 respectively, while for normal LP, ground glass, reticular and honeycombing patterns was 0.89±0.02, 0.70±0.04, 0.72±0.05 and 0.71±0.03, respectively.**

***Index Terms*— computer aided diagnosis, diffuse lung diseases quantification, 3D Co-occurrence, Support Vector Machine classification, MultiDetector CT.**

I. Introduction

COMPUTED tomography (CT) has become the modality of choice for lung imaging. While high resolution CT (HRCT) scan protocols allows visualization of fine lung structures, only a limited portion of lung parenchyma is scanned (approximately 10%) [1]. Multidetector CT (MDCT) allows acquisition of volumentric datasets with almost isotropic voxels, enabling visualization, characterization and quantification of the entire extent of Diffuse Parenchyma Lung Diseases (DPLDs), often presented with a non uniform distribution throughout the lung.

Interpretation of DPLDs is characterized by high inter and intra-observer variability, due to lack of standardized criteria in assessing complex and variable morphological appearance, further complicated by the large amount image data to be reviewed [2].

Computer-Aided Diagnosis (CAD) schemes that automatically detect and quantify radiologic patterns of DPLDs in CT images have been proposed to provide a second opinion to radiologists to improve follow-up management decisions [1], [3]-[8]. These systems generally consist of two stages. The first stage is the segmentation of left and right Lung Field (LF) region by means of grey level-based methods, while the second stage performs differentiation of lung parenchyma into normal and abnormal tissue types, exploiting two-dimensional (2D) texture analysis on/of HRCT datasets [1], [3]-[6] and recently three-dimensional (3D) texture analysis on/of MDCT datasets [7]-[8].

CAD schemes proposed for DPLDs, utilizing HRCT protocols, are targeted to classification of Regions Of Interest (ROIs) into abnormality classes/types by means of 2D textural features such as first order statistics [7], filter based features [1], [4], co- occurrence matrices [7], run lengths [7] and fractal features [7], while classifiers such as Bayesssian [8], neural networks [3] and k-nearest neighbors [1] are exploited.

CAD schemes proposed for DPLDs, utilizing MDCT datasets, exploit 3D texture analysis either for classification of Voxels of Interest (VOIs) into abnormality types [7] or for volumetric detection/quantification of abnormality types by means voxel classification [8]. Xu et al. [7] employed 3D texture features (first order, co-occurrence, run length and fractal features) and compared a Bayessian to a Support Vector Machine (SVM) classifier for classification of VOIs into 5 tissue types. Zavaletta et al. [8] employed LF segmentation by iterative 3D gray level thresholding [9] followed by bronchovascular segmentation as a preprocessing step. Subsequently, histogram signatures originating from a sliding VOI were used in combination with earth's mover's distance, for volumentric quantification of lung parenchyma into 5 tissue types (normal, emphysema, ground glass,

Manuscript received August 13, 2008. This work was supported in part by the Caratheodory Programme (C.180) of the University of Patras

P. Korfiatis is with the Medical Physics Department of University of Patras, Greece (korfp@upatras.gr)

A. Karahaliou is with the Medical Physics Department of University of Patras, Greece (karahaliou.a@med.upatras.gr)

A. Kazantzi is with the Radiology Department of University Hospital of Patras, Greece (akazantzi@yahoo.gr)

C. Kalogeropoulou is with the Radiology Department of University Hospital of Patras, Greece (rat@upatras.gr)

L. Costaridou is with the Medical Physics Department of University of Patras, Greece (phone +302610969111; fax: +30261996113; e-mail: costarid@upatras.gr).

reticular and honey combing).

This paper presents a method for volumetric quantification of Interstitial Pneumonia (IP) patterns, a subset of DPLDs, utilizing a MDCT dataset. Since the accuracy of LF segmentation algorithm influences the performance of the proposed CAD scheme, a 3D LF segmentation algorithm adapted to IP patterns affecting lung borders is employed as a preprocessing step. The 3D LF segmentation algorithm is based on gray level thresholding combined with lung border voxel classification refinement. Following, broncho-vascular structures are segmented utilizing a line enhancement filter and removed from LF, to obtain lung parenchyma (LP) volume. LP volume is subsequently differentiated into normal, ground glass, reticular and honeycombing patterns employing k-means clustering and Support Vector Machine (SVM) voxel classification based on 3D co-occurrence analysis.

The method is differentiated from reported studies towards IP quantification by employing a more elaborate LF segmentation stage, adapted to IP patterns affecting lung borders, while a two stage classification scheme is employed for IP pattern quantification.

II. Materials and Methods

A. Dataset

Clinical cases were acquired from 10 patients subjected to MDCT imaging, diagnosed with IP secondary to connective tissue diseases, radiologically manifested with ground glass, reticular and honeycomb patterns. MDCT scans were obtained with a Multislice (16x) CT (LightSpeed, GE), in the Department of Radiology at the University Hospital of Patras, Greece. Acquisition parameters of tube voltage, tube current and slice thickness were 140 kVp, 300 mA and 0.625 mm, respectively. The image matrix size was 512x512 with average pixel size of 1.25 mm. MDCT scans of 5 patients were used to extract VOIs for training the SVM classifier employed for IP pattern quantification, while MDCT scans of the remaining 5 patients were used for evaluating performance of the proposed method in segmenting LF as well as quantifying normal LP, ground glass, reticular and honeycomb patterns.

B. Data Preprocessing

Lung Field Segmentation

As suggested by Armato and Sensakovic [10], LF segmentation algorithms used as preprocessing step in CAD of lung diseases, should be adapted to the specific abnormalities being quantified. In this work, a two stage 3D LF segmentation technique was employed, adapted to IPs affecting lung borders.

The first stage consists of a previously proposed 3D histogram thresholding LF segmentation algorithm [11]. Specifically, LF highlighting is performed by application of automatically derived gain factors on wavelet magnitude subband image values of two decomposition levels (first and second dyadic scales). Reconstruction from the modified wavelet coefficients provides the lung border highlighted image. Segmentation of the LF volume is achieved by application of minimum error thresholding [12] on lung border highlighted images. A morphological closing operator was applied on the segmented lung border to deal with under-segmentation in the mediastinum area.

However, gray level-based algorithms are insufficient in correctly segmenting LF 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 as a second stage. Specifically, an iterative neighborhood labeling of lung border voxels was performed using an SVM classifier. The SVM classifier assigns a label of LF or surrounding tissue (ST) to a voxel using as inputs local 3D texture features. Specifically, 4 first order statistics (Mean, Standard Deviation, Skewness and Kurtosis) were extracted from a 7x7x7 VOI centered at the voxel being labeled. Voxel labeling is initially applied on each border voxel of the initial LF volume, as provided by the 3D gray level-based algorihtm, and subsequently on its 18-connected neighbors. The initial LF volume is updated by adding voxels labeled as LF and removing voxels labeled as ST (0). The process continues by checking every neighboring voxel of an already labeled one, until the left and right LF volumes stay un-altered. The outermost voxels of corresponding un-altered LF volumes provide the final left and right LF borders. Coordinates of already labeled voxels are stored to avoid double-checking of neighboring voxels during the LF volume updating.

Vessel tree segmentation

Accurate segmentation of vessel tree (broncho-vascular) structures is required to eliminate/reduce false positive errors in final quantification of reticular or honeycomb patterns. Such false positive errors arise due to vessel tree structures radiologic appearance which is similar to that of reticular and honeycomb patterns.

In this work, a vessel tree segmentation algorithm was adopted, recently proposed for vessel tree segmentation in CT angiography images [13]. Specifically, a multiscale line enhancement filter, designed to enhance vessels and vessel bifurcation points, was applied on segmented LF volume, based on the analysis of eigenvalues of the Hessian matrix at multiple scales (σ= 1, 2, …, 12 pixels). An expectation maximization segmentation algorithm was used to segment vessel tree volume by segmenting high response voxels at each level. Finally, reconstruction of segmented vessels components was performed.

The segmented vessel tree volume is removed from the LF volume, resulting in the lung parenchyma (LP) volume.

C. Interstitial pneumonia pattern quantification

The first step in quantification of IP patterns is the

differentiation of normal from abnormal LP. A 2 class *k*-means segmentation algorithm was utilized for this purpose, based on grey level values.

Following, all voxels corresponding to abnormal LP are subjected to supervised 3D texture analysis for quantification of ground glass, reticular and honeycombing patterns utilizing a SVM classification scheme.

3D Co-occurrence features

The Gray Level Co-occurrence Matrix (GLCM) [14] is a well established tool for characterizing the spatial distribution of gray levels in an image (second order statistics). An element at location (*i, j*) of the co-occurrence matrix signifies the joint probability density of the occurrence of gray levels *i* and *j* in a specified direction θ and specified distance *d* from each other. The 3D co-occurrence matrix stores the number of co-occurrences of the pairs of gray levels *i* and *j*, which are separated by a distance *d* (in this study *d*=1 voxel) in a VOI [15]. This 3D method searches for another gray level in 26 directions in multiple planes in constructing the co-occurrence matrix. In this work, the following thirteen 3D co-occurrence matrix features were calculated from each 7x7x7 VOI: 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.

SVM classifier

An SVM voxel classification scheme was used to quantify reticular, ground glass and honeycomb patterns. Vapnik [16] proposed the SVM to find the optimal hyperplane that separates the training data to achieve a minimum expected risk. The basic principles of SVM are the maximal margin of separation and the kernel trick. Considering a two-class pattern classification problem, the SVM first performs a non linear mapping (Φ) from a low-dimensional input space to a higher dimensional feature space via a kernel function

$$K(x_i, x_j) = \Phi^{\mathrm{T}}(x_i)\Phi(x_j) \tag{1}$$

This mapping allows the SVM to achieve better class separation. An SVM can be trained to construct a hyperplane for which the margin of separation is maximized. Maximally separated margins parallel to the hyperplane divide the new feature space into class-specific sub-spaces based on labelled training patterns. The perpendicular distance of the hyperplane from the unknown pattern is used to assign a decision value. There are several choices for the kernel function (*K*). The radial basis

$$K(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{2}$$

where σ is a user-defined parameter, the d^{th} degree polynomial

$$K(x_i, x_j) = (1 + x_i^{\,T} x_j)^d \tag{3}$$

and the sigmoid

$$K(x_i, x_j) = \tanh(k_1 x_i^{\,T} x_j + k_2) \tag{4}$$

kernel functions. The d^{th} degree polynomial kernel is unbounded and can potentially lead to numerical instability, while the sigmoid kernel has two parameters (k_1, k_2) to be tuned [17]. In this work, the radial basis kernel [18] function was utilized. The parameter C, which controls the tradeoff between margin maximization and error minimization, and σ were automatically derived according to radius-margin bound [19].

To classify abnormal LP voxels into ground glass, reticular and honeycomb patterns, a 3-class pattern recognition problem has to be solved. The SVM classifier is a binary classifier, so a strategy, which is called one-against-all classifiers, is applied to build 3-class classifiers utilizing binary classifiers. In this approach, a set of binary classifiers is trained to be able to separate each class from all others. Then each data object is classified to the class for which the largest decision value was determined.

Training Set

The training set of the classifier consisted of 256 cubic VOIs, defined by an expert radiologist, originating from 5 MDCT scans of the dataset. 56 VOIs were selected to represent honeycombing, 100 VOIs were selected to represent Reticular and 100 VOIs were selected to represent ground glass opacities.

D. Performance evaluation

Performance of the proposed method in segmenting LF as well as in quantifying normal LP, ground glass, reticular and honeycomb patterns was evaluated by means of volume overlap on 5 MDCT scans.

In this study, a second radiologist with expertise in CT image interpretation, defined the ground truth by generating manual outlines of LP as well as reticular ground glass and honeycombing patterns. For manual delineation, a tablet (Wacom Intuos3 Tokyo, Japan) was used with an active area of 305×305 mm with 5.080 lpi and accuracy of ±0.25 mm.

The degree of volume *Overlap* between "ground truth" (O) and computer (C) derived borders is defined by:

$$Overlap = \frac{O \cap C}{O \cup C} \tag{5}$$

The value of *Overlap* is bound between zero (no overlap) and one (exact overlap).

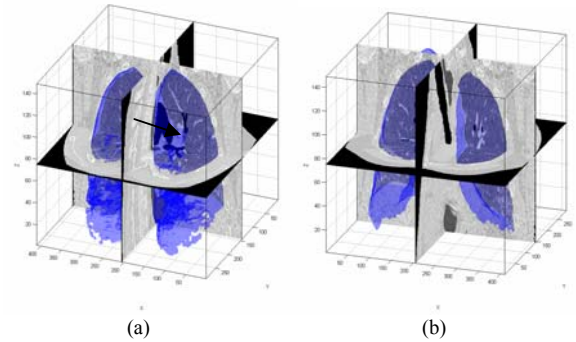
III. Results

Fig. 1 provides an application example of the proposed two stage 3D LF segmentation method. The proposed method

provided accurate LF segmentation results including all the IP affected border areas, as indicated by the arrows.

The volume Overlap separately for left LF and right LF, for the five scans analyzed, were 0.95±0.03 and 0.96±0.02 respectively.

Fig. 2 depicts an application example of vessel tree segmentation (red overlay) corresponding to the same case depicted in Fig. 1.



(a) (b)

Fig. 1. Lung field (LF) segmentation example. Blue overlay corresponds to segmented left and right lung fields. (a) Result of the 3D histogram thresholding with wavelet highlighting. Arrow indicates LF under-segmentation due to IPs affecting border. (b) Result of texture based border refinement step.

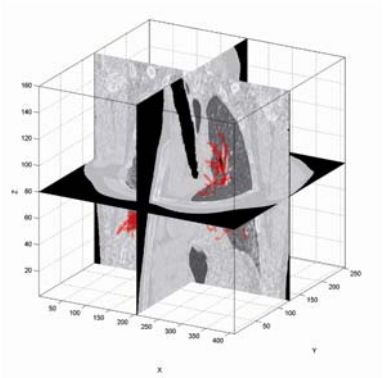


Fig. 2. Vessel tree segmentation example (red overlay).

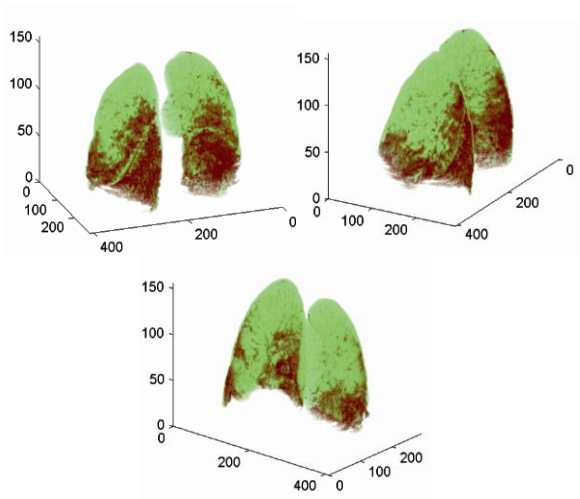


Fig. 3. Application example of k-means clustering for separation of normal (green overlays) from abnormal (red overlays) lung parenchyma at three different views.

Fig. 3 provides an application example of k-means clustering for separation of normal lung parenchyma (air component) from abnormal (IP affected) lung parenchyma at three different views. Normal and abnormal parenchyma voxels are illustrated with green and red colors respectively.

Further differentiation of the abnormal lung parenchyma, of the case depicted in Fig. 3, into ground glass and reticular patterns is provided in Fig. 4. Green overlay corresponds to normal lung parenchyma (same as green overlay of Fig. 3), while yellow and red overlays correspond to ground glass and reticular opacities respectively.

Volume Overlap for normal LP well as for ground glass, reticular and honeycombing patterns were: 0.89±0.02, 0.70±0.04, 0.72±0.05 and 0.71±0.03 respectively.

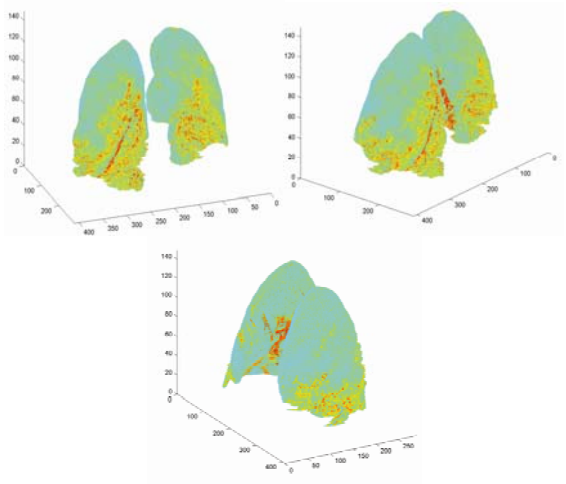


Fig. 4. Quantification example of ground glass and reticular patterns of the abnormal lung parenchyma depicted in Fig. 3. Green overlay corresponds to normal lung parenchyma, while yellow and red overlays correspond to ground glass and reticular opacities respectively.

IV. Discussion

In this study, a system for quantification of DPLDs in MDCT is presented. Use of 3D data enabled application of 3D texture analysis by means of 3D co-occurrence features capable of capturing presence and extent of complex disease patterns.

Quantification of IP patterns is formulated as a three-class pattern recognition problem to classify abnormal lung parenchyma into ground glass, reticular and honeycomb patterns, by means of SVM voxel classification. Abnormal lung parenchyma is provided by application of k-means clustering on previously segmented lung field (LF) volumes.

Accurate LF segmentation is an important initial step in DPLDs quantification, as diseased areas attached to lung borders may not be included in LF (under-segmentation) and not participate in subsequent analysis. In this work, LF segmentation was achieved by 3D gray level thresholding followed by a texture based border refinement step.

The performance of the method in 3D segmentation of LF, normal LP as well as quantification of IP patterns was evaluated by means of volume overlap, providing promising results.

Specifically, a high performance was demonstrated in left and right LF segmentation (0.95 and 0.96 respectively), attributed to the texture-based border refinement step.

The method demonstrated promising results in quantification of IP patterns. However, more accurate differentiation of normal from abnormal lung parenchyma in combination with additional texture features would further improve performance of the method in IP pattern quantification.

V. Conclusion

An automated system for quantification of interstitial lung disease as depicted in MDCT scans is presented. The system is based on three–class SVM classification utilizing 3D co-occurrence features to classify abnormal lung parenchyma voxels into three categories: ground glass, reticular and honeycombing.

Preliminary results are promising, suggesting an accurate and reproducible system. Such systems are expected to assist radiologists in detection and quantification of interstitial DPLDs.

Future extensions of the proposed system should consider expansion of the training and testing datasets, investigation of additional 3D features as well as utilization of a multi-class SVM. The effect of analyzing VOI size in system performance should also be investigated.

Furthermore, performance evaluation should consider shape differentiation metrics, as well as system performance comparison to inter- and intra-observer variability.

References

[1] I. C. Sluimer, M. Prokop, I. Hartmann, and B. van Ginneken, "Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung," *Med. Phys.,* vol. 33, no. 7, pp. 2610-2620, July 2006.

[2] Z. A. Aziz, A. U. Wells, D. M. Hansell, S. J. Copley, S. R. Desai, S. M. Ellis *et al*, "HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation," *Thorax,* vol. 59, no. 6, pp. 506-511, June 2004.

[3] F. Chabat, G.-Z. Yang, and D. M. Hansell, "Obstructive Lung Diseases: Texture Classification for Differentiation at CT," *Radiology,* vol. 228, no. 3, pp. 871-877, Sep. 2003.

[4] H.-U. Kauczor, K. Heitmann, C. P. Heussel,*D.* Marwede T. Uthmann and M. Thelen, "Automatic Detection and Quantification of Ground-Glass Opacities on High-Resolution CT Using Multiple Neural Networks: Comparison with a Density Mask," *Am. J. Roentgenol.,* vol. 175, no. 5, pp. 1329-1334, Nov. 2000.

[5] I. C. Sluimer, P. F. van Waes, M. A. Viergever and B. van Ginneken, "Computer-aided diagnosis in high resolution CT of the lungs," *Med. Phys.,* vol. 30, no. 12, pp. 3081-3090, Dec. 2003.

[6] Y. Uchiyama, S. Katsuragawa, H. Abe, J. Shiraishi, F. Li, Q. Li, *et al.*, "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Med. Phys.,* vol. 30, no. 9, pp. 2440-2454, Aug. 2003.

[7] Y. Xu, E. J. R. van Beek, Y. Hwanjo, J. Guo, G. McLennan and E. Hoffman, "Computer-aided Classification of Interstitial Lung Diseases Via MDCT: 3D Adaptive Multiple Feature Method (3D AMFM)," *Acad. Radiol.,* vol. 13, no. 8, pp. 969-978, Aug. 2006.

[8] V. A. Zavaletta, B. J. Bartholmai, and R. A. Robb, "High Resolution Multidetector CT-Aided Tissue Analysis and Quantification of Lung Fibrosis," *Acad. Radio.y,* vol. 14, no. 7, pp. 772-787, July 2007.

[9] S. Hu, E. A. Hoffman and J. M. Reinhard, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans.Medical Imaging,* vol. 20, no. 6, pp. 490-498, June 2001.

[10] W. F. Sensakovic, S. G. Armato, A. Starkey and P. Caliguiri, "Automated lung segmentation of diseased and artifact-corrupted magnetic resonance sections," *Med. Phy.,* vol. 33, no. 9, pp. 3085-3093, Aug. 2006.

[11] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos*,* C. Kalogeropoulou and L. Costaridou*.*, "Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT," *Br J Radiol,* vol. 80, no. 960, pp. 996-1004, Dec. 2007.

[12] J. Kittler, and J. Illingworth, "Minimum error thresholding," *Pattern Recogn.,* vol. 19, no. 1, pp. 41-47, Jan. 1986.

[13] C. Zhou, H-. P. Chan, B. Shahiner, L. M. Hadjiiski, A. Chughtai, S. Patel *et al*, "Automatic multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications," *Med. Phys.,* vol. 34, no. 12, pp. 4567-4577, Nov. 2007.

[14] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Systems Man and Cybernetics,* vol. 3, no. 6, pp. 610-621, Nov. 1973.

[15] Y. Xu, M. Sonka, G. McLennan, and J. Guo, E.A.Hoffman, "MDCT-based 3-D texture classification of emphysema and early smoking related lung pathologies," *IEEE Trans. Medical Imaging*, vol. 25, no. 4, pp. 464-475, Apr. 2006.

[16] N. V. Vladimir, *The nature of statistical learning theory*: Springer-Verlag New York, Inc., 1995.

[17] S. Chen, S. Zhou, F.-F. Yin, L. B. Marks and S. K. Das, "Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis," *Med. Phys.,* vol. 34, no. 10, pp. 3808-3814, Sep. 2007.

[18] A. Shamsheyeva, A. Shamsheyeva, and A. Sowmya, "The anisotropic Gaussian kernel for SVM classification of HRCT images of the lung," in *Proc. of the Intelligent Sensors, Sensor Networks and Information Processing Conference*, Australia 2004. pp. 439-444

[19] S. S. Keerthi, "Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms," *IEEE Trans. Neural Networks,* vol. 13, no. 5, pp. 1225-1229, Sep. 2002.