Classification of Interstitial Lung Disease Patterns Using Local DCT Features and Random Forest

M. Anthimopoulos, *IEEE Member*, S. Christodoulidis, A. Christe and S. Mougiakakou, *IEEE Member*

Abstract— **Over the last decade, a plethora of computer-aided diagnosis (CAD) systems have been proposed aiming to improve the accuracy of the physicians in the diagnosis of interstitial lung diseases (ILD). In this study, we propose a scheme for the classification of HRCT image patches with ILD abnormalities as a basic component towards the quantification of the various ILD patterns in the lung. The feature extraction method relies on local spectral analysis using a DCT-based filter bank. After convolving the image with the filter bank, q-quantiles are computed for describing the distribution of local frequencies that characterize image texture. Then, the gray-level histogram values of the original image are added forming the final feature vector. The classification of the already described patches is done by a random forest (RF) classifier. The experimental results prove the superior performance and efficiency of the proposed approach compared against the state-of-the-art.**

I. Introduction

Interstitial lung diseases (ILDs) constitute a heterogeneous group of more than 200 chronic lung disorders characterized by scarring and/or inflammation of lung parenchyma that cause respiratory failure [1]. The diagnosis of ILD is usually based on the assessment of thoracic computed tomography (CT). However due to the lack of strict clinical guidelines and the resemblance between the different ILD findings, the problem of radiological ILD diagnosis is time consuming and requires extensive experience, a fact that reduces significantly the diagnostic accuracy of the radiologists and increases the inter- and intra- observer variability up to 50% [2]. Optimal treatment and prognosis of ILDs depend on accurate diagnosis. Misdiagnosis may lead to serious and life-threatening complications [1]. Therefore, additional invasive procedures are often required like transbronchial or even surgical biopsies, increasing both the risk and the cost of the diagnosis. To this end, a lot of research effort has been put into the development of computer aided diagnosis (CAD) systems able to improve the diagnostic accuracy.

II. Related Work

A typical CAD system for ILD consists of two major stages: (a) lung segmentation and (b) ILD pattern quantification. In the first stage the pulmonary parenchyma is extracted whereas in the second, the existing ILD pathologies are identified and quantified by classifying local image patches. In this study, we will focus on ILD quantification stage which has attracted considerable attention over the last decade. A variety of mainly texture feature sets have been proposed for the description of patches and different classification techniques for recognizing the described area.

The first proposed approaches employed already established statistical tools for the description of lung tissue such as first-order statistics, gray level co-occurrence matrices (GLCM), run-length matrices (RLM) and fractal analysis [3]-[6]. The concatenation of them results in the so-called adaptive multiple feature method (AMFM) as proposed by Uppaluri et al. [3]. Uchiyama et al. [7] introduced the use of geometrical measures based on the morphological top-hat transforms while Sluimer et al. [8] used histogram moments after filtering with multi-scale Gaussian, Laplacian and Gaussian derivative filters. Vo et al. [9] proposed a similar description model with additional wavelet and contourlet features. Sorensen et al. [10] used local binary patterns (LBP) combined with a gray level histogram.

Recently, several studies adopted unsupervised feature extraction techniques based on Bag of Features (BoF) [11]-[13] and sparse representation (SR) models [14]-[16]. In these methods, a set of texture atoms, namely textons are identified by k-means and k-SVD, respectively, and every local structure in the image is represented by the closest texton or a linear combination of the entire set. Image description is usually done by estimating the frequency distribution of the various existing textons. For the description of textons in the BoF models Gangeh et al. [11] proposed the use of raw pixel values, Rodriguez et al. [12] used a wavelet transform on the differences of Gaussians, while Xu et al. [13] computed the moments on pixel values and local Hessian eigenvalues. The same features were also used within a SR approach in [14] by the same team. Vo et al. [15] proposed a SR-based method using wavelet-contourlet features, whereas Zhang et al. [16] built their SR model on the normalized intensity values and then used a joint RS-intensity histogram. Li et al. [16] utilized a Gaussian Restricted Boltzmann Machine (GRBM) for the learning of 16 filters for each of the 3 considered scales and then used the sum of the filter responses as features.

After the description stage, a feature vector is created and a machine learning classifier is usually employed in order to assign to each patch an ILD pattern label. Many different classification methods have been used such as linear discriminant (LD) [6], [8] and Bayesian [3], [5] classifiers, the k-Nearest Neighbors (kNN) [8], [10], [12], artificial neural networks (ANN) [4], [7], and support vector machines (SVM) with linear [14], [17] or Radial Basis function (RBF) [8], [11], [13] kernels. Vo et al. proposed the use of a

S. Mougiakakou, M. Anthimopoulos and S. Christodoulidis are with ARTORG Center for Biomedical Engineering Research, University of Bern, Switzerland

S. Christodoulidis is with the Dept. of Electr. & Comput. Eng., Aristotle Univ. of Thessaloniki, Greece

A. Christe is with the Department of Diagnostic, Interventional and Pediatric Radiology, Bern University Hospital "Inselspital", Switzerland

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

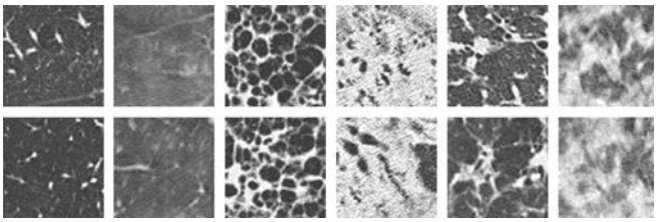

Figure 1. Two example CT representations of each ILD tissue pathologies. From left to right: healthy, GGO, honeycombing, consolidation, reticulation, reticulation/GGO.

multiple kernel learning classifier (m-MKL) in [9], while in [15] they used the minimum reconstruction error as a classification criterion after reconstructing the patch using class-specific dictionaries.

III. METHODS

In this study, we propose a method for the description and classification of HRCT image patches with ILD pathologies based on local 2D discrete cosine transform (DCT) transform and random forest (RF) classification. The considered classes are six including healthy tissue and five ILD patterns, namely ground glass opacity (GGO), consolidation, reticulation, honeycombing and the combination of reticulation with GGO. Examples of the aforementioned lung tissue patterns are presented in Fig. 1.

A. Feature Extraction

Many different filters have been proposed for the description of local image structures such as edges, ridges and blobs by emphasizing specific frequencies in the two-dimensional spectrum. In this work we used a filter bank consisting of the 2D-DCT orthogonal basis functions, covering the entire frequency spectrum of a local area. The DCT describes a discrete signal as a linear combination of cosine functions oscillating at different frequencies. The two-dimensional DCT of an $N\times N$ image $I(x,y)$ is defined as:

$$F(u,v) = a(u)a(v)\cdot \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} I(x,y)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (1)$$

where $u, v = 0, 1, ..., N-1$ and $a(x) = \begin{cases} \sqrt{1/N} & \text{if} \quad x = 0 \\ \sqrt{2/N} & \text{if} \quad x \neq 0 \end{cases}$

Each of the DCT coefficients in $F(u,v)$ corresponds to one 2D frequency of the DCT basis and represents its contribution in the linear combination of frequencies that results in the original image. Fig. 2 presents the basis of the 5×5 DCT with the frequencies increasing from left to right and from top to bottom.

The convolution of an image with the $N\times N$ DCT basis functions produces an equal number of images, each of them describing the contribution of a local frequency in the original image. The distributions of the filter bank responses carry information regarding the local spectral content of the image and could be used for characterizing its texture. In order to capture this information we propose the use of a relatively small number of quantiles. Quantiles are points

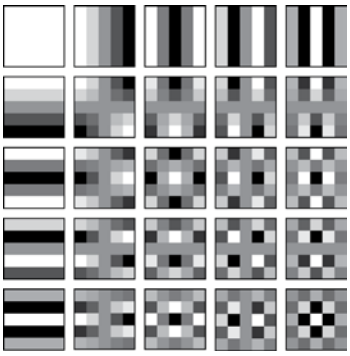

Figure 2. The basis functions of the 5x5 DCT.

taken at regular intervals from the cumulative distribution function (CDF) of a random variable. Even a few quantiles are usually able to describe accurately the distribution of a random variable while providing robustness against noise.

If the number of considered intervals is q, q-1 values are defined, often referred to as q-quantiles. In practice, CDF is approximated by the cumulative histogram and the q-quantiles for every filtered image are computed and concatenated into a feature vector. The histogram of the original image is also added since the grayscale distribution of CT images expresses fundamental physical properties of the tissue. The motivation behind using histogram values for the original image but the quantiles for the spectral analysis lies in the fact that noise is more prominent in high frequencies. The final size of the feature vector is $(N\times N)*Q + B$, where N is the size of the local DCT transform, Q is the number of quantiles considered and B is the number of bins of the gray-level histogram. The feature extraction procedure is depicted in Fig. 3.

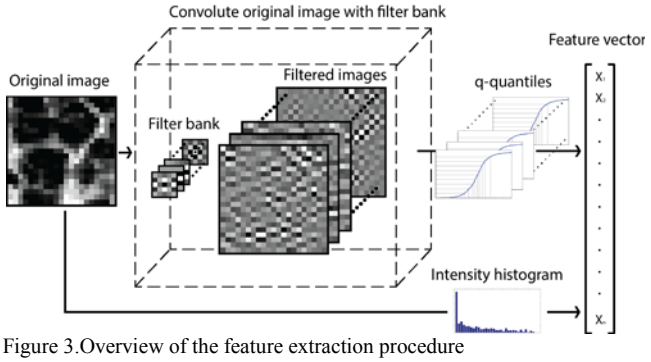

Figure 3.Overview of the feature extraction procedure

B. Classification

After the feature extraction procedure, image patches are represented by feature vectors which are fed to a random forest classifier in order to be classified. RFs are ensemble classifiers proposed by Breiman [18]. A RF is a combination of decision trees with each tree depending on the values of a randomly sampled feature vector. To classify a new input vector, each tree "votes" for a class and the forest chooses the class having the majority of votes over all the trees in the forest. RFs have been used successfully in numerous machine learning applications yielding classification performances at least comparable to SVM and ANN while being much faster,

especially in the prediction phase. Moreover, RFs can handle very large numbers of input variables and they are fully parallelizable and easily implemented. However, to the best of our knowledge, they have never been used before for CT-image lung tissue classification.

IV. Experimental Setup and Results

A. Data

For the needs of the present study, the publicly available TALISMAN database was used [19]. The database consists of 113 HRCT scans with an average of 25 slices per scan and 512×512 pixels per slice. It also includes annotated regions of interest (ROIs) for 17 different pathological patterns along with clinical parameters from patients with histologically proven diagnoses of ILDs. For the current study, six lung patterns were considered and re-annotated by experienced radiologists of the Bern university hospital - Inselspital. Based on the annotated ROIs for the healthy and the five selected ILD patterns, a dataset of nearly 2500 ILD image patches was created with size equal to 21×21pixels. The patches are entirely included in the lung field and they have an overlap with the corresponding ROI of at least 90% while the maximum overlap among the patches is 33%. Table I presents an overview of the database.

TABLE I. Database Overview

Tissue category	# Cases	# ROIs	# Patches
Normal	15	103	506
GGO	16	152	506
Honeycombing	10	28	343
Consolidation	12	34	136
Reticulation	18	61	506
Reticulation/GGO	41	177	506

B. Evaluation

The evaluation procedure for the classification of the ILD patches is based on a 5-fold cross-validation scheme. The database is randomly split into 5 subsets and each of them is used for testing after having trained with the rest. The performance is assessed in terms of the average F-score over the different classes:

$$F_{avg} = \frac{1}{M}\sum_{c=1}^{M} F_c \tag{2}$$

where M is the number of classes and F_c is the F-score for class c, defined as:

$$F_c = 2\frac{precision_c \cdot recall_c}{precision_c + recall_c} \tag{3}$$

With

$$recall_c = \frac{samples\_correctly\_classified\_as\_c}{samples\_of\_class\_c} \tag{4}$$

$$precision_c = \frac{samples\_correctly\_classified\_as\_c}{samples\_classified\_as\_c} \tag{5}$$

C. Results

The best values for the parameters were identified by a trial and error procedure while cross-validating on the whole dataset. Specifically, the optimal filter size was found equal to 5×5 resulting in 25 filters. For the q-quantiles, q was set to 10 giving 9 values to which the max and min values were also added. The gray-level histogram of the original images was clustered to 32 bins producing 307 features, in total. Adding more quantiles or histogram values increased the computational complexity without improving the performance. For the RF classifier, 40 trees were considered sufficient and the number of variables at each node was set equal to the square root of the total number of features. The experiments were carried out using MATLAB on an Intel i5-2500 @3.30 GHz CPU with 8GB of RAM.

TABLE II. Comparison of Filter Banks for the Feature Extraction

Filter Bank	Number of filters	F_{avg} (%)
Leung-Malik [20]	48	76.6
Schmid [21]	13	78.7
Gabor [22]	36	79.2
MR8 [23]	8	81.3
DCT	25	89.0

In order to justify the choice of the proposed DCT filters for the feature extraction, we compared with some of the most popular filter banks found in literature keeping the rest of the system as proposed. Table II includes the results of the experiment that prove the advanced performance of the chosen filters due to their ability to capture information from the entire spectrum.

TABLE III. Comparison of Methods for the Description of Filter Responses

Classifier	Values per image	F_{avg} (%)
Moments	4	86.9
Histogram	11	87.5
Quantiles	11	89.0

The next experiment compares different ways to capture the relevant information from the filter responses for the feature extraction. Table III provides results for using (i) four histogram moments (mean, standard deviation, skewness, and kurtosis), (ii) histogram values and (iii) the proposed quantiles. In all cases the original image is described by its histogram. The results prove the robustness of quantiles in describing frequency distributions.

TABLE IV. Comparison of Classifiers Using the Proposed Feature Set

Classifier	F_{avg} (%)	Training time (sec)	Testing time (sec)
kNN	80.0	0.043	0.295
SVM	83.1	14.641	0.526
ANN	84.3	1.451	0.013
RF	89.0	4.064	0.004

After the experiments regarding the feature extraction we conducted a comparison of four popular classifiers trained on the proposed DCT-based feature set. The overall classification performance along with the average execution time for the training and testing of each classifier are summarized in Table IV. An extensive search was performed to identify the best parameters for every classifier; the k parameter for kNN was set to 11, ANN was implemented as

a two-layer feed-forward network with 43 hidden and 6 output neurons while for SVM the RBF kernel was used with *C*=4 and *gamma* = 16. From the results, one can confirm the superiority of the RF which yielded the best performance while keeping a low training time and the lowest by far time of prediction.

Furthermore, Table V provides a comparison of the proposed approach against four state-of-the-art methods as implemented by the authors.

TABLE V. COMPARISON OF THE PROPOSED METHOD AGAINST FOUR STATE-OF-THE-ART SYSTEMS

Method	**Features**	**Classifier**	F_{avg}**(%)**
Sluimer [8]	Moments after filter bank	SVM-RBF	65.7
Uppaluri [3]	GL Moments, GLCM, RLM	Bayesian	77.5
Gangeh [11]	Raw pixel textons	SVM-RBF	78.9
Sorensen [10]	LBP + GL histogram	kNN	82.7
Proposed	**Quantiles of local DCT + GL histogram**	**RF**	**89.0**

Finally, Fig. 4 presents the confusion matrix for the proposed method. The relatively high misclassification rate between the combined GGO/reticulation and the individual GGO and reticulation patterns could be justified by the fact that the former constitutes an overlap of the latter. Moreover, honeycombing is misclassified as reticulation with a rate of 8% due to their common fibrotic nature while 7% of GGO is classified as normal since they both share similar texture with just a slight difference in the average intensity.

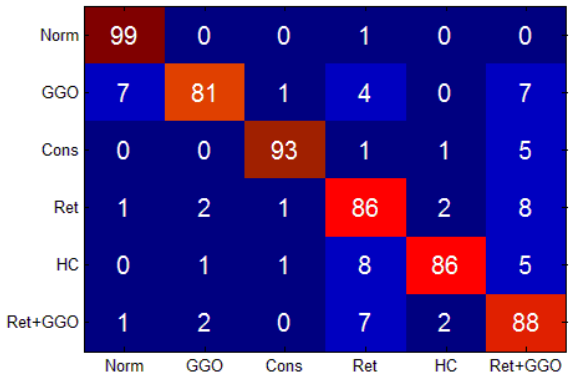

Figure 4. The confusion matrix of the proposed classification scheme. The entry in the i^{th} row and j^{th} column corresponds to the percentage of images from class i that was classified as class j. Norm: normal tissue; GGO: ground glass opacity; Cons: consolidation; Ret: reticulation, HC: honeycombing; Ret+GGO: combination of reticulation and ground glass opacity.

V. CONCLUSION

In this paper, we proposed a method for classifying lung tissue with ILD pathologies based on local spectral analysis and random forest classification. For the spectral analysis, a filter bank consisting of the 5×5 DCT basis functions was used. The distribution of the filtered images was described by the 10-quantiles (deciles) together with the max and min values while a 32-bin gray-level histogram of the original image was also considered for the feature set. The highly discriminative feature set was combined with a RF classifier which was used for the first time on the specific application. The experimental results proved the superior performance and efficiency of the proposed method achieving the best results with an average F-score equal to 89%. For future work we will investigate the extension of the proposed 2D fixed-scale filter bank to multiple scales and 3 dimensions.

REFERENCES

[1] A.U. Wells, Managing diagnostic procedures in idiopathic pulmonary fibrosis. *Eur Respir Rev*. 2013;22(128):158-62.

[2] I. Sluimer, et al., "Computer analysis of computed tomography scans of the lung: a survey," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 385-405, April 2006.

[3] R. Uppaluri et al., "Computer recognition of regional lung disease patterns," *Am J Respir Crit Care Med*., vol. 160, no. 2, pp. 648-54, 1999.

[4] KR. Heitmann et al., "Automatic detection of ground glass opacities on lung HRCT using multiple neural networks," *Eur Radiol*., vol. 7, no. 9, pp. 1463-72, 1997.

[5] F. Chabat et al. "Obstructive lung diseases: texture classification for differentiation at CT," *Radiology*, vol. 228, no. 3, pp. 871-7, 2003.

[6] S. Delorme et al., "Usual interstitial pneumonia. Quantitative assessment of high resolution computed tomography findings by computer-assisted texture-based image analysis," *Invest Radiol*, vol. 32, no. 9, pp. 566-74, 1997.

[7] Y. Uchiyama et al., "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Med Phys*., vol. 30, no. 9, pp. 2440-54, 2003.

[8] I. Sluimer et al., "Computer-aided diagnosis in high resolution CT of the lungs." *Med Phys.,* vol. 30, no.12, pp.3081-90, 2003.

[9] K.T. Vo et al., "Multiple kernel learning for classification of diffuse lung disease using HRCT lung images," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS 2010*, pp.3085,3088.

[10] L. Sorensen et al., "Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns," *IEEE Transactions on Medical Imaging,* , vol.29, no.2, pp.559,569, Feb. 2010

[11] M. Gangeh et al., "A texton-based approach for the classification of lung parenchyma in ct images", *Med Image Comput Comput Assist Interv*. Vol. 13(Pt 3), pp.595-602, 2010

[12] A. Foncubierta-Rodríguez et al., "Using Multiscale Visual Words for Lung Texture Classification and Retrieval", *Medical Content-Based Retrieval for Clinical Decision Support, Lecture Notes in Computer Science,* vol. 7075, pp 69-79, 2012

[13] R. Xu et al., "Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach." *Med Image Comput Comput Assist Interv.*, vol.14(Pt 3), pp.183-90, 2011

[14] W. Zhao et al., "Classification of diffuse lung diseases patterns by a sparse representation based method on HRCT images," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS 2013*, pp.5457-5460, 2013.

[15] K.T. Vo et al., "Multiscale sparse representation of HRCT lung images for diffuse lung disease classification," In *Int'l Conf. on Image Processing (ICIP)*, pp.441-444, 2012

[16] M. Zhang et al. "Pulmonary Emphysema Classification based on an Improved Texton Learning Model by Sparse Representation", Proc. SPIE, vol. 8670, 2013.

[17] Q. Li et al., "Lung image patch classification with automatic feature learning," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS 2013*, pp.6079-6082, 2013

[18] L. Breiman, "Random Forests", *Machine Learning, vol.*45, no.1, pp. 5-32, 2001

[19] A. Depeursinge et al., "Building a reference multimedia database for interstitial lung diseases", *Comput Med Imaging Graph*., vol.36, no.3, pp.227-38, 2012

[20] T. Leung et al. "Representing and recognizing the visual appearance of materials using three-dimensional textons", *International Journal of Computer Vision,* vol.43, no.1, pp.29-44, 2001

[21] C. Schmid, "Constructing models for content-based image retrieval," in *Proc. CVPR*, vol.2, pp.II-39,II-45, 2001

[22] I. Fogel et al., "Gabor filters as texture discriminator", *Biological Cybernetic,* vol.61, no.2, 1989.

[23] M. Varma et al. "Classifying images of materials: Achieving viewpoint and illumination independence" in *Proc. ECCV*, vol. 200, no.23, pp.255–271