2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Texture-based classification of lung disease patterns in chronic hypersensitivity pneumonitis and comparison to clinical outcomes

F. Pennati, L. Aliboni, A. Antoniazza, D. Beretta, O. Dias, B. G. Baldi, M. Sawamura, R. C. Chate, C.R.R. De Carvalho, A. Albuquerque, A. Aliverti

Abstract— **Computer-aided detection algorithms applied to CT lung imaging have the potential to objectively quantify pulmonary pathology. We aim to develop an automatic classification method based on textural features able to classify healthy and pathological patterns on CT lung images and to quantify the extent of each disease pattern in a group of patients with chronic hypersensitivity pneumonitis (cHP), in comparison to pulmonary function tests (PFTs).**

27 cHP patients were scanned via high resolution CT (HRCT) at full-inspiration. Regions of interest (ROIs) were extracted and labeled as normal (NOR), ground glass opacity (GGO), reticulation (RET), consolidation (C), honeycombing (HB) and air trapping (AT). For each ROI, statistical, morphological and fractal parameters were computed. For automatic classification, we compared two classification methods (Bayesian and Support Vector Machine) and three ROI sizes. The classifier was therefore applied to the overall CT images and the extent of each class was calculated and compared to PFTs. Better classification accuracy was found for the Bayesian classifier and the 16x16 ROI size: 92.1±2.7%. The extent of GGO, HB and NOR significantly correlated with forced vital capacity (FVC) and the extent of NOR with carbon monoxide diffusing capacity (DLCO).

Clinical Relevance— **Texture analysis can differentiate and objectively quantify pathological classes in the lung parenchyma and may represent a quantitative diagnostic tool in cHP.**

I. Introduction

Different forms of interstitial lung disease (ILD) have been described, all characterized by the presence of inflammation and altered lung interstitium [1]. The importance of a correct diagnosis of ILDs is crucial to determine the correct therapy. Differential diagnosis is difficult, as ILDs have rather similar clinical manifestations. HRCT is the protocol of choice for differential diagnosis and is based on the extent and the distribution of the various ILD textural patterns, including reticulation, honeycombing, ground glass opacity and consolidation[2]. This is considered a difficult task for radiologists, because of the complexity and the various appearance of the disease patterns, leading to high intra- and inter-observer variability[3]. Invasive procedures, such as bronchoalveolar lavage and histological confirmation, are required in ambiguous cases. A reliable, computer-aided diagnosis could assist the radiologist, improving diagnostic accuracy and avoiding surgical lung biopsies for some patients. Automated classification systems have been developed to objectively classify and quantify the extent of different diseased patterns[3]–[10]. One approach is to extract relevant texture and shape features from regions of interest (ROIs) and use supervised machine learning techniques to classify the ROI as healthy or pathological. Different textural features have been investigated, including first- and second-order gray level statistics, fractal dimension[3], [4], wavelet transform [7], [11] and filtering techniques[12]. We aim at developing an automated classification method based on textural and shape features, to evaluate the extent of regional disease patterns in a group of patients with chronic hypersensitivity pneumonitis (cHP). Also, we aim at comparing quantification results to clinical measures.

II. Materials and Methods

A. Data

27 patients with cHP were included. cHP diagnosis was based on CT findings, known antigenic exposure, exclusion of other possible diagnoses, and compatible histology. Forced vital capacity (FVC) and DLco were measured and expressed as percent predicted. Patients underwent CT scanning at suspended full inspiration on a Brilliance 16P scanner (Philips Medical Systems, Andover, MA). Scanner settings: tube voltage, 120-140 kVp; tube current, 30 mAs; matrix 512x512; slice thickness, 1 mm, smooth reconstruction filter (B30f). Study protocol was approved by the local research ethics committee and all patients signed informed consent.

B. ROI selection

Using software MIPAV (Medical Image Processing, Analysis and Visualization, v. 7.2.0; http://mipav.cit.nih. gov), an experienced pneumonologist and an experienced radiologist, manually selected and classified ROIs into air trapping (AT), normal (NOR), ground glass opacity (GGO), reticulation (RET), consolidation (CON), and honeycombing (HB). Ambiguous regions presenting more than one pattern, or whose classification was uncertain, were neglected. To standardize ROI shape and size, squared ROIs centered in the centroid of the manually drawn regions, sized 8x8, 16x16 and 32x32 pixels, were extracted. Before ROIs extraction, edgementation to merge the neighboring segments based on the difference in their gray-scale values was applied [4].

B. Textural and shape feature extraction

Each ROI was characterized by computing textural and shape features (Table 1): gray-level parameters (first and second order statistics), morphological parameters, fractal parameters and wavelet parameters.

1) Gray-level parameters represent the texture based on the distribution and the relationship of gray level values in the

F. Pennati, L. Aliboni, A. Antoniazza, D. Beretta and A. Aliverti are with the Dipartimento di Elettronica, Informatica e Bioingegneria, Politecnico di Milano, Milan, ITALY (corresponding author F. Pennati, phone: +39-0223999026, e-mail: francesca.pennati@polimi.it).

O. Dias, B. G. Baldi, M. Sawamura, R. C. Chate, C.R.R. De Carvalho, A. Albuquerque are with Instituto do Coracao (InCor), Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo - São Paulo, BRAZIL.

image. First-order parameters provide statistical properties of the intensity histogram of the image. In this study, we calculated: mean, mode, standard deviation, skewness and kurtosis. Second-order parameters, which characterize the spatial relationship of gray levels in the image, were determined starting from gray level co-occurrence (GLCM) and run-length matrices. The GLCM is computed by counting the number of occurrences of the pair of gray level i and j, which are at a distance d apart, across a defined direction. In this study, three GLCMs corresponding to three different directions (0°, 45°, 90°) and one distance d=1 pixel, were computed for each ROI. From each GLCM, four features were derived: contrast, energy, homogeneity and entropy. The run-length matrix searches for runs of pixels of the same grey value across a given direction, with coarse and fine textures defined respectively as large and small numbers in the run. Short-run emphasis and the long-run emphasis were calculated.

2) Morphological parameters were included to provide information on the shape of the opacities, based on morphological filtering operations [13] and on Minkowsky functionals [14].

The white top-hat transform (defined by subtracting the opening of an image from the original image), was applied to enhance nodular and linear (reticular) components, using a square of 4x4 pixels as a structuring element. The filtered image was binarized with an arbitrary threshold equal to 175 (gray level value), to remove small noise [13]. The resulting binary structures were split according to their circularity: structures with a circularity >=0.7 were considered as candidates for being nodular components (nodular mask), whereas structures with a circularity <0.7 were considered as candidates for being linear components (linear mask). The average pixel value of the original image multiplied respectively by the nodular and the linear mask was defined as a measure of nodular and linear components in each ROI.

The morphological black top-hat transform (defined by subtracting the original image from the opening of the image), was applied to highlight the multilocular component (honeycombing pattern), using a square of 7x7 pixels as a structuring element. The mean, standard deviation and entropy were calculated from this image.

Minkowsky functionals (MF) characterize binary images in terms of shape (geometry) and connectivity (topology). Defining a set of thresholds, the gray-level images of the lung pattern can be represented as a set of binary images [14]. For each binary image, three Minkowsky functionals are defined: area (number of white pixels), perimeter (sum of the perimeters of the different elements in the image), and Euler number (the number of elements in the image minus their holes) [15]. We applied two thresholds to binarize the images: <-910 HU, to characterize honeycombing and law attenuation area, and >-230 HU, to identify consolidation.

3) Fractal parameters. Fractal is a term used to represent patterns that possess self-similarity across scales or levels of magnification and have been successfully applied to pulmonary imaging, for identifying low attenuation area [16]. In this study, we evaluated if fractals were able to characterize patterns other than low attenuation area. Thus, we applied four gray-level ranges: (-1000, -960) HU for low attenuation areas [16]; (-1000, -910) HU for honeycombing; (-900, -870) HU for normal; and (-850, -750) HU for ground glass opacity. Fractal dimension was computed using boxes with size equal to a power of 2, so the boxes size ranged from 1 to 2^p, where p was the smallest integer such that ROIs dimension was smaller or equal to 2^P. N (number of boxes of dimension R needed to cover the non-zero values of the ROI), and R (boxes dimension), were graphed on a log-log plot, showing a linear relation. The fractal dimension is the mean value of the slopes in the log-log plot.

4) Wavelet transform parameters. Wavelet transform iteratively decomposes an image into several components based on the frequency content and orientation. In this study, we applied a level 2 decomposition via Haar wavelet transform. On the details of the transformed image in the horizontal, vertical, and diagonal directions, energy was computed.

TABLE I. SUMMARY OF FEATURES THAT REPRESENT EACH ROI

Category	***Features***
Histogram	mean, mode, standard deviation, skewness and kurtosis
Second order parameters	co-occurance matrix (energy, contrast, homogeneity, entropy) , run-length matrix (SRE, LRE)
White top-hat transform	linear & nodular components
White top-hat transform	mean, standard deviation, entropy
Minkowsky functionals	MF. Area, MF. Perimeter, MF. Euler
Fractal parameters	Fractal dimension
Wavelet transform	Energy of details

C. Feature subset selection

Parameters' selection is intended to select the best subset of texture descriptors that retain the maximum information content with respect to the original set of variables, improving classification performance and speeding up the computational time. A stepwise forward selection was used. The selection was applied 10 times to obtain 10 parameters subsets; from these subsets, the parameters which appear at least once were selected. Nevertheless, considering that the final application of the classifier is on the entire image and that the classifier must deal with regions at the interface between classes, we also performed an experimental parameters' selection to exclude the parameters which were highly influenced by intermediate areas. The final subset included: histogram mode, second-order gray-level parameters, MF.Euler and fractal dimensions.

D. Classification

Two classifiers were applied for comparison: a naive Bayesian classifier and a support vector machine (SVM) classifier. To classify unknown ROIs, models were created from known labeled data in the training phase, to find the optimal classifier parameters that model the best boundaries between the classes.

Overall, the dataset included 349 ROIs: 51 air trapping (AT), 23 consolidation (CON), 79 ground-glass opacity (GGO), 23 honeycombing (HB), 83 reticulation (RET) and 90 normal (NOR). The dataset was divided in training and test set. The test set consisted of 10 observations for each class, randomly selected, thus the overall number of observations in the test set was equal to 60. The training set consisted of the remaining observations (initial training set) on which the Synthetic Minority Over-sampling TEchnique (SMOTE)

algorithm was applied (final training test)[17], [18]. This technique creates artificial data based on the feature space similarities between existing minority samples. This technique is able to generate synthetic instances rather than replicate minority class instances; therefore, it can avoid the over-fitting problem. The total number of observations in the training set after the application of the SMOTE was 480, balanced between the classes. On the final training set, a model was trained, and its performances were tested on the relative test set. To estimate prediction accuracy, 100 iterations were performed, randomly selecting the test and the training sets at each iteration (Fig. 1). Classifiers' performance was assessed through sensitivity (proportion of observations correctly classified for each group), specificity (proportion of observations classified in a specific group that really belonged to that group), and accuracy (proportion of observations correctly classified over the entire test set).

D. Volume processing

The trained system was applied to HRCT images of the whole lung in each patient. Lung segmentation was performed using a density-based approach, which selects the automatic threshold to separate low-density tissue from the surrounding chest wall[19]. To make the segmentation more reliable to the anterior and posterior junction, a wavelet pre-processing is applied before thresholding[20]. Central airways and central vascular structures, reconstructed respectively using the 3D confidence connected region growing algorithm and a tubular structure enhancement filter[21], were subtracted from the parenchyma mask. Each class was quantified on 8 slices, equally spaced from top diaphragm to aortic arch. On each slice, a square window scrolled on each lung pixel, and the centered pixel of the window was classified. The volume fraction of each regional disease pattern was calculated.

E. Statistical analysis

Data are reported as mean±standard deviation unless otherwise stated. Correlation between each class extent and lung function parameters was evaluated via Spearman correlation. Statistical analysis was performed in IBM SPSS® Statistics (IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0.0.1 Armonk, NY).

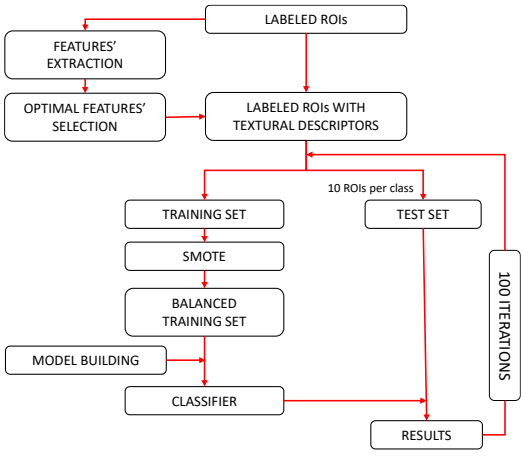

Figure 1. Classification system.

III. RESULTS

A. Evaluation of classifier and ROI size

Table II shows the sensitivity and the specificity of the Bayesian and the SVM classifiers, with an overall accuracy respectively equal to 92.1±2.7% and 91.4±3.0%. Although the classifiers' performances were similar, differing just in the sensitivities of HB and NOR patterns, we chose the Bayesian classifier for the lower computational cost.

Table III shows the performance of the Bayesian classifier based on ROI size 8x8, 16x16 and 32x32, with an overall accuracy respectively equal to 79±10%, 92±3% and 86±4%. ROI sized 16x16 shows the higher accuracy and the higher sensitivity and specificity for all classes.

B. Parameters' selection

The results obtained with the subset of parameters experimentally selected were compared to the results obtained with the subset selected via the forward selection method. The accuracy of the subset obtained with the forward selection was higher than the experimental subset, respectively 95.0±2.5% and 92.1 ± 2.7%. The sensitivity and precision were comparable between the two subsets, except for the NOR class, respectively 87.9 ±9.4% and 76.2 ± 12.4%.

C. Evaluation of patients and correlation with PFTs

Representative images from four patients with the corresponding classification results are shown in Fig. 2. FVC correlated with the extent of GGO (ρ=-0.54, $p<0.01$), HB (ρ=-0.48, $p<0.05$) and NOR (ρ=0.58 $p<0.01$). CON and AT were difficult to be interpreted because of their low extent, respectively 0.1±0.3% and 0.5±0.9%. DLCO negatively correlated with NOR (ρ=-0.48, $p<0.05$).

TABLE II. SENSITIVITY AND SPECIFICITY COMPARISON OF BAYESIAN AND SVM CLASSIFICATION

	Sensitivity (%)		***Specificity (%)***	
	Bayesian	***SVM***	***Bayesian***	***SVM***
AT	94±8	94±7	86±8	93±7
NOR	76±12	87±11	89±10	83±10
GGO	96±6	90±10	89±8	89±8
RET	93±8	96±6	98±4	91±7
C	100±0	100±2	100±2	99±3
HB	95±6	83±9	94±7	99±4

Sensitivity, and specificity for each texture feature and each classifier. The means and the standard deviations of the 100 iterations are reported. (ROI size 16x16).

TABLE III. PERFORMANCE OF THE CLASSIFIER BASED ON THE ROI SIZE

	Sensitivity (%)			***Specificity (%)***		
	8	**16**	**32**	***8***	**16**	**32**
AT	94±6	94±8	91±10	89±8	86±8	88±8
NOR	73±16	76±12	80±10	87±10	89±10	83±11
GGO	66±26	96±6	94±8	86±16	89±8	83±9
RET	66±34	93±8	81±10	90±10	98±4	86±10
C	96±18	100±0	79±10	96±5	100±2	96±7
HB	76±17	95±6	91±8	57±17	94±7	87±10

Sensitivity, and specificity for each texture feature and each ROI dimension. The means and the standard deviations of the 100 iterations are reported. (Bayesian classifier)

IV. DISCUSSION

The present study developed a CT-based texture analysis method of the lung parenchyma, to evaluate the extent of regional disease patterns in cHP. We implemented common statistical, morphological and fractal features, which have been successfully applied to discriminate diseased patterns[3], [13], [14], [16], trying to optimize the classifier performances. In particular, Minkowsky functionals were determined by combining information derived from the gray-level distribution (gray-level thresholds), and shape hypothesis (area thresholds). Moreover, fractal dimensions were computed on images binarized with different thresholds to characterize patterns other than low attenuation areas, as originally proposed[16]. Also, the SMOTE algorithm was applied to balance the classes. The Bayesian and the SVM classification methods, had comparable accuracy, with the Bayesian classifier characterized by a higher negative predictive value for the healthy pattern. The 16x16 ROI, have higher sensitivity and specificity for all the patterns except for the air trapping, similarly to Chang et al[22]. Generally, a lower ROI produced a classification with a higher variability, while a larger ROI produced more homogeneous regions. Also, a smaller ROI was able to detect patterns characterized by a specific attenuation level, as air trapping, while it was not able to detect morphological patterns like HB, that requires a larger dimension. The results of the quantitative CT analysis were also correlated to pulmonary function tests, showing its potential as a quantitative tool in cHP.

Future work will be oriented to increase the dataset and reducing the impact of the artificial data created by the SMOTE technique. Also, as manual ROI selection may make the results biased, training in other ILDs dataset will be performed. A 3D analysis of the lung volume would probably improve patient evaluation, but with higher computational costs[9]. Moreover, just inspiratory images were considered, but it would be of interest to evaluate if the expiratory images could provide additional information.

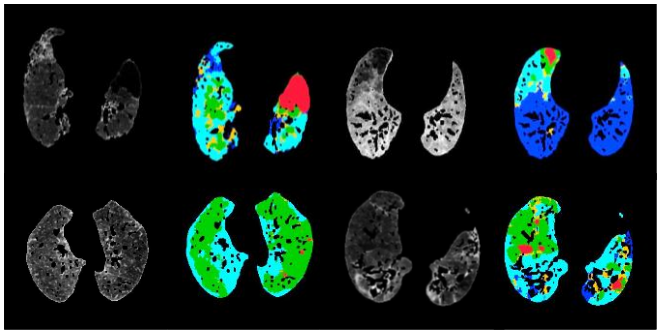

Figure 2. Representative lung CT images from four cHP patients, with the corresponding classification results. Each pixel was color-coded: normal, green; low attenuation area, red; ground-glass opacity, cyan; reticular opacity, blue, honeycombing, yellow; and consolidation, dark blue.

REFERENCES

[1] K. C. Meyer, "Diagnosis and management of interstitial lung disease," *Transl. Respir. Med.*, 2014.

[2] N. L. Muller, "Clinical value of high-resolution CT in chronic diffuse lung disease," *American Journal of Roentgenology*. 1991.

[3] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan, "Computer recognition of regional lung disease patterns," *Am. J. Respir. Crit. Care Med.*, 1999.

[4] R. Uppaluri, T. Mitsa, M. Sonka, E. A. Hoffman, and G. McLennan, "Quantification of pulmonary emphysema from lung computed tomography images," *Am. J. Respir. Crit. Care Med.*, 1997.

[5] L. Aliboni *et al.*, "Quantitative CT Analysis in Chronic Hypersensitivity Pneumonitis: A Convolutional Neural Network Approach," *Acad. Radiol.*, no. November, 2020.

[6] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[7] A. Depeursinge *et al.*, "Optimized steerable wavelets for texture analysis of lung tissue in 3-D CT: Classification of usual interstitial pneumonia," in *Proceedings - International Symposium on Biomedical Imaging*, 2015.

[8] 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.*, 2006.

[9] Y. Xu, E. J. R. van Beek, Y. Hwanjo, J. Guo, G. McLennan, and E. A. 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, 2006.

[10] S. Delorme, M. A. Keller-Reichenbecher, I. Zuna, W. Schlegel, and G. Van Kaick, "Usual interstitial pneumonia: Quantitative assessment of high- resolution computed tomography findings by computer-assisted texture-based image analysis," *Invest. Radiol.*, 1997.

[11] K. T. Vo and A. Sowmya, "Multiple kernel learning for classification of diffuse lung disease using HRCT lung images," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, 2010.

[12] 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.*, 2003.

[13] Y. Uchiyama *et al.*, "Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography," *Med. Phys.*, vol. 30, no. 9, pp. 2440–2454, 2003.

[14] M. B. Huber, M. B. Nagarajan, G. Leinsinger, R. Eibel, L. A. Ray, and A. Wismüller, "Performance of topological texture features to classify fibrotic interstitial lung disease patterns," *Med. Phys.*, vol. 38, no. 4, pp. 2035–2044, 2011.

[15] K. R. Mecke and H. Wagner, "Euler characteristic and related measures for random geometric sets," *J. Stat. Phys.*, 1991.

[16] M. Mishima *et al.*, "Complexity of terminal airspace geometry assessed by lung computed tomography in normal subjects and patients with chronic obstructive pulmonary disease," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 16, pp. 8829–8834, 1999.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, 2002.

[18] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, 2009.

[19] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Trans. Med. Imaging*, vol. 20, pp. 490–498, 2001.

[20] F. Pennati, C. Salito, and A. Aliverti, "Registration of lung CT images acquired in different respiratory ranges with 4DCT and HRCT," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015.

[21] J. T., P. F., L. B., and S. Z., "Enhancement of Vascular Structures in 3D and 2D Angiographic Images," *IEEE Trans. Med. Imaging*, 2016.

[22] Y. Chang, J. Lim, N. Kim, J. B. Seo, and D. A. Lynch, "A support vector machine classifier reduces interscanner variation in the HRCT classification of regional disease pattern in diffuse lung disease: Comparison to a Bayesian classifier," *Med. Phys.*, vol. 40, no. 5, 2013.