

Research Article

Enhanced Classification of Interstitial Lung Disease Patterns in HRCT Images Using Differential Lacunarity

Verónica Vasconcelos,^{1,2} João Barroso,^{1,3} Luis Marques,² and José Silvestre Silva⁴

¹INESC TEC, Campus da FEUP, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

²Coimbra Institute of Engineering, Polytechnic Institute of Coimbra, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

³School of Science and Technology, University of Trás-os-Montes e Alto Douro, Apartado 1013, 5001-801 Vila Real, Portugal

⁴Military Academy Research Center, Avenida Conde Castro Guimarães, 2720-113 Amadora, Portugal

Correspondence should be addressed to Verónica Vasconcelos; veronica@isec.pt

Received 11 September 2015; Accepted 18 November 2015

Academic Editor: Yukihisa Takayama

Copyright © 2015 Verónica Vasconcelos et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The analysis and interpretation of high-resolution computed tomography (HRCT) images of the chest in the presence of interstitial lung disease (ILD) is a time-consuming task which requires experience. In this paper, a computer-aided diagnosis (CAD) scheme is proposed to assist radiologists in the differentiation of lung patterns associated with ILD and healthy lung parenchyma. Regions of interest were described by a set of texture attributes extracted using differential lacunarity (DLac) and classical methods of statistical texture analysis. The proposed strategy to compute DLac allowed a multiscale texture analysis, while maintaining sensitivity to small details. Support Vector Machines were employed to distinguish between lung patterns. Training and model selection were performed over a stratified 10-fold cross-validation (CV). Dimensional reduction was made based on stepwise regression (F -test, p value < 0.01) during CV. An accuracy of $95.8 \pm 2.2\%$ in the differentiation of normal lung pattern from ILD patterns and an overall accuracy of $94.5 \pm 2.1\%$ in a multiclass scenario revealed the potential of the proposed CAD in clinical practice. Experimental results showed that the performance of the CAD was improved by combining multiscale DLac with classical statistical texture analysis.

1. Introduction

Interstitial lung disease (ILD) is a common name for a heterogeneous group of complex disorders affecting lung parenchyma. The ILD affects similar lung regions and has identical clinical, radiological, and functional tests which hinder the differential diagnosis. However, ILD subtypes have different prognoses and treatments, so a correct diagnosis is essential [1]. High-resolution computed tomography (HRCT) imaging of the chest can offer such good image quality that it has become essential in the detection, diagnosis, and follow-up of ILD [2]. HRCT images of patients affected with ILD have specific patterns whose distribution and visual content analysis is particularly relevant in elaborating an accurate diagnosis [3].

Multidetector row computed tomography (CT) scanners generate a huge volume of data that must be visually examined by radiologists. This task is very time-consuming

and requires experience, especially in the presence of ILD. Computer-aided diagnosis (CAD) for ILD is seen as a necessary tool to reduce interobserver and intraobserver variations, as well as to improve diagnostic accuracy by assisting radiologists in the detection, characterization, and quantification of pathological regions [3–13].

In this paper, a CAD scheme is presented allowing for a classification of regions of interest (ROIs), from HRCT images, in four classes of lung patterns: normal (NOR), ground glass (GG), honeycombing (HC), and emphysema (EMP). A scenario of binary differentiation, NOR class versus pathological class, is also considered. A generic flowchart of the proposed approach is shown in Figure 1. Classical statistical methods were used to extract and quantify texture information. The first-order (FO) analysis, the Spatial Gray Level Dependence Method (SGLDM), and the Gray Level Run-Length Method (GLRLM) allowed the estimate of statistical properties of individual pixel values and of the

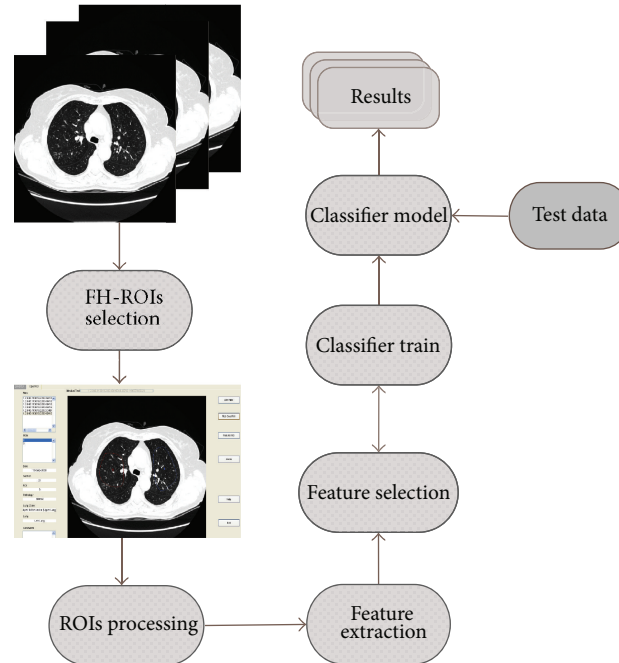


FIGURE 1: The proposed CAD scheme.

spatial interaction between two or more pixel values. These methods have frequently been used in texture analysis of medical images, namely, in the description of ILD patterns [4, 8–13]. Given the heterogeneity of lung parenchyma in healthy subjects or in the presence of pathologies, a multiscale texture analysis was proposed using differential lacunarity (DLac). Lacunarity has been successfully used in analyzing medical images of different organs or structures, acquired by different types of equipment. In [14, 15], fractal lacunarity analysis was applied to lumbar vertebra magnetic resonance images in order to extract relevant parameters, allowing for differentiation among three types of trabecular bone structure, from female subjects with different age and physiopathological status. In [16], lacunarity was combined with mean fractal dimension to differentiate between aggressive and nonaggressive malignant lung tumors, in sequence of contrast-enhanced CT images. An 83.3% accuracy can be valuable information in the choice of the appropriate treatment procedure. In [17], lacunarity analysis was applied for discriminating endoscopic images, obtained through a wireless capsule endoscopy technique, related to a common interstitial disease: ulceration. A promising classification accuracy of over 97% was obtained. In [18], lacunarity was applied to HRCT images of the chest to differentiate between normal and emphysematous regions of lung parenchyma. The preliminary results showed the potential of the proposed lacunarity features.

After a feature selection procedure, the obtained features were used to classify each ROI through a Support Vector Machines (SVM) algorithm. This learning algorithm has its origin in statistical learning theory and structural risk minimization [19, 20]. It emerged as an efficient technique for solving classification problems. A comparative study

between SVM and other popular classifiers was performed by Meyer et al. [21]. The results highlight that SVM classifiers are among the best. In [5], five common classifiers were compared according to their ability to differentiate six lung tissue patterns in HRCT images. The results showed that SVM provides the best trade-off between the error rate and the capacity for generalization, an important aspect to take into consideration given the diversity of pulmonary patterns.

2. Materials and Methods

2.1. Texture Analysis. Texture is a major component in the interpretation of HRCT images in the presence of ILD. The most difficult aspect of texture analysis is to define a set of meaningful features that describe the texture associated with different lung patterns. Each ROI of $M \times N$ pixels was represented by a set of m features extracted using the methods described in the sections below.

2.1.1. First-Order Statistics Analysis. The CT attenuation of each ROI was described through FO statistical features extracted from ROI normalized histogram. Considering that L is the number of gray levels used in ROI quantization, the normalized histogram $h(z_i)$, $0 \leq i < L$, gives the probability of observing the gray level z_i in the ROI. From $h(z_i)$, six statistics features were computed: the mean, variance, skewness, kurtosis, energy, and entropy [22].

2.1.2. Spatial Gray Level Dependence Method. The method of texture analysis proposed by Haralick et al. [23] describes the spatial dependence of gray level distribution between neighboring pixels. In the SGLDM, the second-order joint conditional probability distribution $p(i, j \mid dx, dy)$ can be

estimated for a defined length and along a defined direction given by offsets in x and y direction: dx and dy . So, $p(i, j | dx, dy)$ is the probability that two pixels at a distance given by (dx, dy) have the gray levels i and j . The function $p(i, j | dx, dy)$ is defined as follows:

$$p(i, j | dx, dy) = \frac{\sum_{k=1}^M \sum_{q=1}^N \delta(i, \text{ROI}(k, q)) \cdot \delta(j, \text{ROI}(k + dx, q + dy))}{T(dx, dy)}, \quad (1)$$

where $\delta(i, j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$

$\text{ROI}(k, q)$ is the intensity at pixel (k, q) and $T(dx, dy)$ is the total number of pixels pairs belonging to the ROI in the length and direction given by (dx, dy) . The functions $p(i, j | dx, dy)$ can be written in matrix form $\Omega(dx, dy) = p(i, j | dx, dy)$, $0 \leq i, j < L$, where L is the maximum gray level of the ROI. For each pair (dx, dy) , a different matrix $\Omega(dx, dy)$ can be computed. Often, each matrix $\Omega(dx, dy)$ is calculated taking into account a given offset and its opposite, giving rise to symmetrical matrices. In this study, from each matrix, six textural measures were extracted: angular second moment, entropy, inverse difference moment, correlation, contrast, and variance [23, 24].

2.1.3. Gray Level Run-Length Method. Run-length primitives were computed by the GLRLM [25]. A run-length primitive is a consecutive and collinear set of pixels with the same gray level. These primitives can be characterized by their length, direction, and gray level. Each chosen direction gives rise to a run-length matrix $\Psi(\theta)$ whose elements represent the number of runs with gray level intensity a and length r , along the direction θ :

$$\Psi(\theta) = M(a, r | \theta), \quad 0 \leq a < L, \quad 0 < r \leq N_r, \quad (2)$$

where L is the number of gray levels and N_r is the possible maximum run-length in ROI along θ direction. From each run-length matrix $\Psi(\theta)$, eleven features were extracted, listed, and described in [24–27].

2.1.4. Lacunarity Analysis. Most of the textures and natural surfaces tend to have a fractal dimension (FD) that can be seen as a measure of irregularity [28]. However, different textures and natural surfaces can share identical FD. In order to differentiate these types of fractal patterns, Mandelbrot [29] proposed lacunarity, a complementary measure of FD that describes the texture of a fractal or their deviation from translational invariance [30]. More recent studies introduced lacunarity analysis as a technique that can be used to describe general spatial patterns, regardless of whether it is a fractal [31]. By using lacunarity, it is possible to distinguish the texture of spatial patterns through the analysis of their distribution gap sizes, at different scales.

Due to the extensive range of gray levels used in CT images acquisition, an appropriate algorithm to calculate lacunarity is that proposed in [32], called DLac. It is based

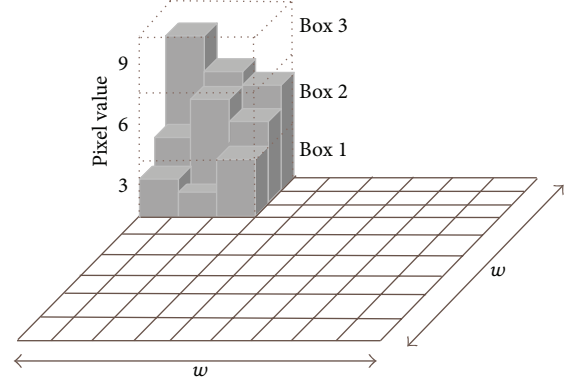


FIGURE 2: Differential box counting algorithm. A moving window of 9×9 pixels and a gliding box of 3×3 pixels are used to compute the box mass. A column of 3 cubic boxes is generated. The differential height of the column is $n(1, 1) = 3 - 1 - 1 = 1$.

on the gliding box [33] and the differential box counting algorithms [34].

According to DLac algorithm, the ROI is divided into overlapped windows of size $w \times w$ pixels, which scans the entire ROI, and a box of size $r \times r$ pixels, which scans each window ($r < w$). The box is placed on the left corner of the window w and a column of accumulated cubes of size $r \times r \times r$ is used to cover the ROI intensity surface in the box place (Figure 2). A sequential number is assigned to each cubic box, from bottom to top. Considering that the maximum and minimum pixel values lie in the cubic box v and u , respectively, the differential height of the column is given by $n(i, j) = v - u - 1$, where (i, j) is the box position. The box mass M of the window w , at specific coordinates, is obtained gliding the box inside the entire window w :

$$M = \sum_{i,j} n(i, j). \quad (3)$$

Considering $n(M, r)$, the number of windows w with box mass M calculated through a box r , the respective probability function $Q(M, r)$ is obtained by dividing $n(M, r)$ by the total number of windows. The DLac of the ROI for a box r , given a window w , is defined as follows:

$$\Lambda(r) = \frac{\sum_M M^2 Q(M, r)}{[\sum_M M Q(M, r)]^2}. \quad (4)$$

2.2. Feature Selection. After performing feature extraction, it is important to proceed with the selection of the most informative features. The resulting set of optimal features improves the classifier performance, while providing a reduction of the general data, as well as a better understanding of the data. The feature selection methods can be divided into two main groups: the filter methods and the wrapper methods. In the filter methods, the features are ordered based on a relevance index. In the wrapper methods, the process of feature selection involves the predictor. In these methods, subsets of features are scored during the learning machine training according to their predictive power [35].

In this work, the reduction of dimensionality was performed using the filter method stepwise regression [36]. In this systematic method, terms are added or removed from the multilinear model based on their statistically significant p value of F -statistics. The method begins with an initial model to which terms that have p values less than an entrance tolerance are added, step by step, and the model terms with p values greater than an exit tolerance are removed from the model.

2.3. Support Vector Machines. The reduced number of parameters that need to be tuned as well as the good trade-off between the error rate and the capability of generalization of SVM classifier algorithm was decisive for its choice in the classification of lung patterns [5, 21].

The SVM strategy, known as the kernel trick, is to map the input data space into a higher dimension feature space, via a nonlinear function kernel $\Phi: \mathcal{R}^m \rightarrow \mathcal{F}$, where separability between classes is improved. The distance between the nearest points of the two classes (*margin*) is maximized, creating an optimal separating hyperplane (OSH).

Considering the training data $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, l$, $\mathbf{x}_i \in \mathcal{R}^m$, $y_i \in \{+1, -1\}$, each instance \mathbf{x}_i is characterized by a vector of m features (or attributes) and is associated with a class +1 or -1. The SVM machine learning solves the following quadratic optimization problem:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i (\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (5)$$

where \mathbf{w} is a normal vector to OSH and b is the bias. In *hard margin* SVM all the examples have to stay outside the margin and be well classified. However, in real datasets, it is necessary to deal with outliers that can be inside the margin or on the wrong side of the classification boundary. The solution proposed in [37], as the *soft margin* SVM, is to introduce constraint slack variables ξ_i in the optimization problem. Ideally, these variables should be zero or have small values. So, to minimize the contribution of the slack variables, a penalty term C is added to the objective function (5). This parameter is a trade-off between the maximization of the margin and the minimization of training errors. For a test example \mathbf{x} , the decision function is given by

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \Phi(\mathbf{x}) + b). \quad (6)$$

There are several functions kernels $K(\mathbf{x}, \mathbf{x}_i) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)$ which can be selected to solve nonlinear problems. In this work, the Gaussian Radial Basis Function (RBF) was used: $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2)/(2\sigma^2)$. This function only has one parameter (σ) that has to be tuned during the classifier training and model selection.

As the standard SVM is a binary classifier, several methods were developed to extend SVM to an n -class problem. Typically, these methods are based on combinations of binary classifiers such as one-versus-all and one-versus-one. In the

one-versus-all approach, n binary classifiers are trained. For example, the model of the n th classifier is trained using the training instances of the n -class as positive and all the instances of the other classes as negative. To classify a new instance, all the n classifiers are run on this instance. The assigned class corresponds to the classifier which returns the largest distance from the separating hyperplane. In the one-versus-one approach $n(n-1)/2$ classifiers are trained in a pairwise methodology, where each takes one class as positive and the other class as negative. To classify a new example, each classifier is run and a count is assigned to each class selected by the classifier. The new instance is classified as belonging to the class which obtains the greatest number of wins, such as in a winner-takes-all voting scheme [38].


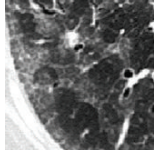
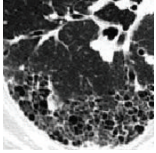
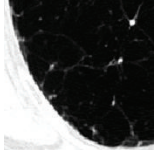
2.4. Dataset. The dataset \mathcal{D} used in this work was acquired in Radiology Department of Coimbra Hospital and University Centre, Coimbra, Portugal. It contains examples of representative regions associated with GG, HC, EMP, and NOR lung patterns, obtained from the daily practice of the hospital. The examples were acquired from subjects that agreed with the use of their images for research purposes by a written consent.

A user friendly software was developed to visualize CT exams, to outline freehand ROIs (FH-ROIs), and to label and to characterize each FH-ROI [39]. HRCT scans were acquired using multidetector row CT LightSpeed VCT 64, from General Electric Healthcare, with an average voxel size of $0.7 \times 0.7 \times 1.3 \text{ mm}^3$, without contrast agent. Each image was stored in a matrix of 512×512 pixels, with 16-bit gray level, using DICOM standard. Each image was displayed using a lung window with a centre in -700 Hounsfield Units (HU) and a width of 1500 HU. From CT images of 57 subjects (#29 female; #28 male) with an average age of 61 ± 16 years, radiologists outlined FH-ROIs from patients in different stages of disease.

The area and shape of each FH-ROI depend on the size and localization of the lung patterns. No more than one FH-ROI was selected from each side of the lungs. The lung region of each FH-ROI was sampled and covered with contiguous, nonoverlapping ROIs of 40×40 pixels [24]. Each FH-ROI was sequentially numbered and each ROI holds the reference of the FH-ROI from where it was extracted. For example, the FH-ROI x Sy corresponds to ROI y extracted from FH-ROI x . Only the ROIs one hundred percent inside the FH-ROI boundary were considered in the train and test of the classifier; all the other ROIs were discarded. For example, in Figure 3 only the ROIs 8, 13, 14, and 18 respect the constraint. Table 1 resumes the dataset used to train and evaluate the proposed CAD system.

2.5. Model Selection and Performance Evaluation. The dataset \mathcal{D} (#1261 ROIs) was divided into a training set and a testing set in a proportion of 2/3-1/3, respectively. The samples were randomly selected using a holdout strategy with stratification, which ensures mutually exclusive partitions where the class proportions are roughly the same as those in the original dataset \mathcal{D} [40]. The holdout procedure was based on FH-ROIs in order to ensure that ROIs extracted from the same FH-ROI are placed on only one of the sets, train or test.

TABLE 1: Dataset used to train and evaluate the CAD system.

Class	Normal	Ground glass	Honeycombing	Emphysema
Visual aspect				
# of patients	16	20	7	14
# of freehand ROIs	87	166	72	92
# of ROIs	253	396	217	395

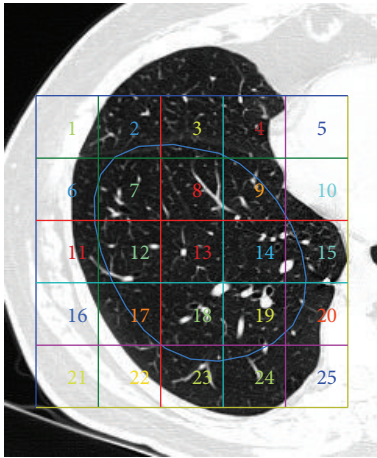


FIGURE 3: Example of FH-ROI and the grid that allows the extraction of ROIs. Only ROIs of one hundred percent inside FH-ROI boundary were kept.

During SVM training, a search was carried out to find optimal parameters to create the classifier model. In the case of the selected RBF kernel function, the parameters that have to be tuned were σ and C , the regularization parameter that corresponds to a penalty over the training errors. The search for the optimal parameters was done using a grid search methodology in the hyperparameter space. So, for every point of the hyperparameter space, a k -fold stratified cross-validation (CV) was performed, with $k = 10$ [40, 41]. The train set was randomly split into k mutually exclusive folds F_1, F_2, \dots, F_k , with approximately the same proportion of each class as in \mathcal{D} . During CV, the classifier was trained and tested k times. In each iteration, it was trained on $k - 1$ folds and tested in the remaining fold F_t , with $t \in \{1, 2, \dots, k\}$. The average of the k -fold accuracy corresponds to CV accuracy. To avoid the model overfitting, the feature selection procedure was included in CV loop [35]. The parameters and features that allow the best CV accuracy were selected and a fine grid search was carried out around the selected parameters, for refinement. The final classifier model was built using all the training set, the selected features, and the optimal parameters previously found. The obtained model was evaluated in the test set, which was not used during classifier training.

TABLE 2: Generic contingency table for n -class scenario.

Actual	Predicted		
	Class 1	\dots	Class n
Class 1	a_{11}	\dots	a_{1n}
\vdots	\vdots	\ddots	\vdots
Class n	a_{n1}	\dots	a_{nn}

The performance evaluation of the classifier was performed based on a contingency table, as exemplified in Table 2 for n -class. Each matrix element has two indices; the first one corresponds to actual disease, while the second one corresponds to predicted disease. The elements of the main diagonal have equal indices representing correct classifications. All the other elements of the matrix correspond to incorrect classifications. For example, a_{31} means that a patient with disease 3 was misclassified as having disease 1. In the case of a binary classification, there are only normal and pathological classes, in a one-versus-all configuration.

After classification, the contingency table was filled with the obtained results. From these values, it is possible to compute a set of metrics allowing for the evaluation of the classifier performance. A common performance evaluation is overall accuracy, which measures the proportion of correctly classified instances for all the classes. Sensitivity of class i measures the fraction of actual positive instances of that class that are correctly classified, while precision measures the correctness of the predictions for class i . Specificity of class i measures the fraction of actual negative instances of class i that are correctly classified. These metrics can be computed by the following expressions [42]:

$$\begin{aligned}
 \text{Sensitivity}(i) &= \frac{a_{ii}}{\sum_j a_{ij}}, \\
 \text{Precision}(i) &= \frac{a_{ii}}{\sum_j a_{ji}}, \\
 \text{Specificity}(i) &= \frac{1 - \sum_j a_{ij} - \sum_{j \neq i} a_{ji}}{1 - \sum_j a_{ij}}, \\
 \text{Overall Accuracy} &= \frac{\sum_{i=j} a_{ij}}{\sum_i \sum_j a_{ij}}.
 \end{aligned} \tag{7}$$

2.6. Feature Settings. Each ROI was characterized by a feature vector extracted using FO, SGLDM, GLRLM, and DLac.

The ROIs were quantized to 32 gray levels before the extraction of FO, SGLDM, and GLRLM features. The minimum and maximum HU values were calculated for all ROIs of \mathcal{D} and each ROI was quantized according to this range. In SGLDM and GLRLM, the directions $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ were considered. In GLDM, the distance between neighboring pixels was $d = 1$, in the four directions.

A multiscale approach was required due to the high variability of the appearance of lung patterns, even for the same pattern. The selected approach to calculate DLac allows a texture analysis at different scales by changing the value of w , for a box size r . The size of the window w determines the coarseness of the scale. The size of r should be relatively small in order to maintain sensitivity to small details present in the neighboring areas. Equation (8) illustrates the proposed approach to extract DLac features:

$$\Lambda(w, r) = \frac{\sum_M M^2 Q(M, r, w)}{[\sum_M M Q(M, r, w)]^2}. \quad (8)$$

DLac was computed for every box-window combination, subject to the condition $r < w$, in order to evaluate the DLac features that better differentiate the lung patterns. A DLac curve $\Lambda(w, r = \text{const})$ can be obtained by keeping the size of the gliding box r constant and by changing the size of the window w . To assure a common referential, DLac values were normalized in relation to the DLac value corresponding to the smallest window w : $\Lambda(w_{\min}, r = \text{const})$ [43]. In order to take advantage of the extensive scale used in CT images, the curves of normalized DLac were computed using Hounsfield scale $[-1000 \text{UH}; +1000 \text{UH}]$.

3. Results and Discussion

Two scenarios were considered in order to evaluate the potential of the proposed CAD and the importance of DLac features in the CAD performance improvement. In the first approach, the differentiation between normal and ILD patterns was considered. The next step was the differentiation of the four classes. In both cases, the feature vector was obtained using two different sets. Set 1 includes the features from FO + SGLDM + GLRLM. Set 2 also englobes DLac features.

The DLac features were extracted from DLac normalized curves. Various experiments have been conducted computing DLac curves for every box-window size for $r = [2-34]$ pixels and $w = [3-35]$ pixels. The ability to differentiate the four classes was evaluated. The best results were obtained for DLac normalized curves for $r = 4$ pixels and $w = [5-35]$ pixels.

Figure 4 shows the average of normalized DLac curves for patterns of all the dataset \mathcal{D} . The results show that the DLac normalized curves are able to distinguish between lung patterns, being suitable to extract informative features.

The multiclass classification was performed using one-versus-one implementation [44]. In the case of the RBF kernel function, the parameters optimization was performed for the pair (C, σ) . First, the parameters were evaluated using a coarse grid for $C = 2^{-5}, 2^{-4.5}, \dots, 2^{15}$ and

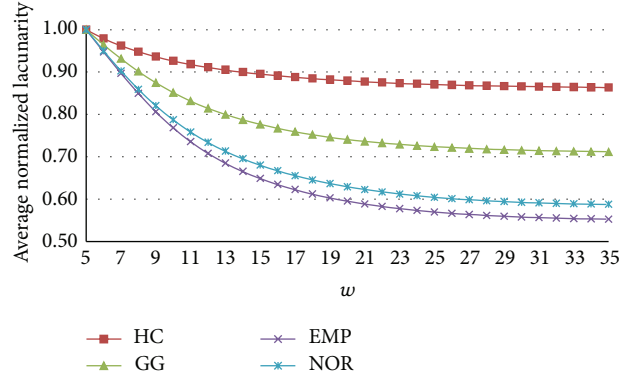


FIGURE 4: Averaged normalized DLac curves obtained for $r = 4$ pixels and $w = [5-35]$ pixels.

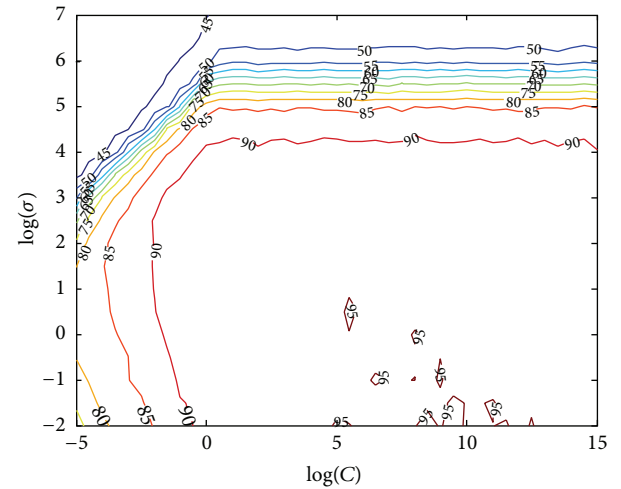


FIGURE 5: Example of 10-fold CV accuracy (%) obtained along the hyperparameter space for finding optimal parameters (C, σ) . Results were obtained using Set 1, for the binary classification scenario.

$\sigma = 2^{-2}, 2^{-1.5}, \dots, 2^7$. Figure 5 depicts a graphic of contours of CV accuracy, obtained over a 10-fold CV using features of Set 1, for the binary classification scenario. A heuristic analysis of these curves provides a clear understanding of the influence of the parameters in the classifier performance, as well as clues to reduce search space. The results showed the importance of fine tuning the SVM parameters during the classifier training phase to achieve an optimized model. After some experiments, the search grid was reduced to $C = 2^3, 2^{3.5}, \dots, 2^{13}$ and $\sigma = 2^{-2}, 2^{-1.5}, \dots, 2^1$. For every coordinate of the hyperparameter space, a k -fold CV was performed, with $k = 10$. In each of the k iterations feature selection was performed (F -test, p value < 0.01) in the $k - 1$ training folds. If the coordinates $(2^c, 2^\sigma)$ generate the best CV accuracy, a finer search was performed around these values with a step of 0.25 upward and downward.

After the classifier training, the selected model was evaluated in the testing set. The training and testing of the classifier were repeated over fifty iterations. So, the training and evaluation of the classifier were performed in fifty different sets.

TABLE 3: Mean (SD) accuracy, sensitivity, precision, and specificity using Set 1 and Set 2, for the binary classification (normal versus pathological). Values in percentage, obtained for 50 iterations.

	Set 1	Set 2
Accuracy	94.4 (2.0)	95.8 (2.2)
Sensitivity	96.7 (1.2)	97.9 (1.1)
Precision	96.0 (2.1)	96.9 (2.1)
Specificity	84.8 (8.6)	88.1 (8.0)

TABLE 4: Mean (SD) of class-specific sensitivity, precision, and specificity using Set 1, for the multiclass classification. Values in percentage, obtained for 50 iterations.

	Classes			
	NOR	GG	HC	EMP
Sensitivity	87.2 (4.6)	92.5 (2.4)	89.4 (4.6)	96.9 (1.7)
Precision	89.6 (3.8)	84.7 (4.5)	93.5 (2.6)	99.8 (0.5)
Specificity	97.3 (1.1)	93.4 (2.3)	98.6 (4.6)	99.9 (0.2)

The presented metrics are the average of the results obtained over all the iterations.

In the binary classification scenario, the ROIs with normal pattern (#253) were considered as negative instances and the other ones as positive instances (#1008). In Table 3, the mean and standard deviation (SD) of overall accuracy, sensitivity, precision, and specificity are shown. The classifier performance obtained using features of Set 1 was $94.4 \pm 2.0\%$ for accuracy, $96.7 \pm 1.2\%$ for sensitivity, $96.0 \pm 2.1\%$ for precision, and $84.8 \pm 8.6\%$ for specificity. Using Set 2, the results increased to $95.8 \pm 2.2\%$ for accuracy, $97.9 \pm 1.1\%$ for sensitivity, $96.9 \pm 2.1\%$ for precision, and $88.1 \pm 8.0\%$ for specificity. High sensitivity and small SD values showed that the proposed CAD has the ability to signal the presence of abnormal patterns using both sets of features; that is, the number of false negatives is low. The integration of DLac features has primarily increased the specificity value, 3.3% on average, reducing the number of false positives. However, the SD value remains high (8.0). The correct classification of NOR class instances is not easy due the high variability of healthy lung tissue.

The classifier performance in the multiclass scenario was also improved using Set 2 (Tables 4 and 5). The overall accuracy increased from $91.9 \pm 1.9\%$ to $94.5 \pm 2.1\%$. Moreover, the class-specific metrics for NOR, GG, and HC improved in a higher or lower percentage. For class EMP sensitivity slightly increased from 96.9% to 97.3%, the precision and the specificity maintained excellent values of 99.9%. Sensitivity of NOR class was the metric that most improved with a mean increase of about 5.3%, changing from 87.2% to 92.5%. In the case of NOR class these results mean that class-specific false negatives decreased; that is, the number of instances of NOR class that were categorized as pathological instances is smaller. In a clinical environment, this means that fewer patients are subjected to the stress of unnecessary additional medical exams. NOR class-specific precision and specificity improved from 89.6% to 92.3% and from 97.3% to 97.9%, respectively. These results mean that the number of

TABLE 5: Mean (SD) of class-specific sensitivity, precision, and specificity using Set 2, for the multiclass classification. Values in percentage, obtained for 50 iterations.

	Classes			
	NOR	GG	HC	EMP
Sensitivity	92.5 (3.8)	96.7 (3.0)	92.3 (4.0)	97.3 (1.8)
Precision	92.3 (3.4)	88.9 (3.9)	97.5 (2.5)	99.9 (0.3)
Specificity	97.9 (1.1)	95.2 (1.8)	99.5 (4.0)	99.9 (0.1)

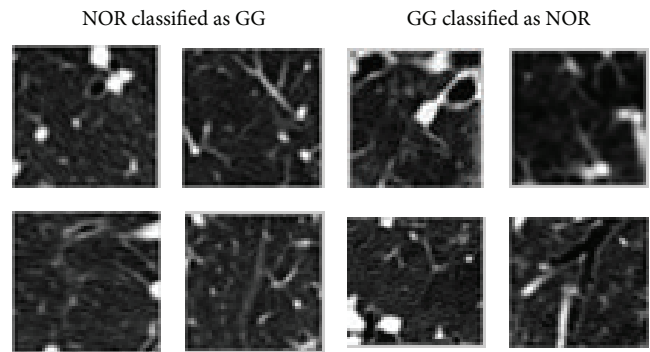


FIGURE 6: Examples of misclassified ROIs between GG and NOR classes.

false positives for NOR class, that is, pathological instances classified as normal, decreased with Set 2. This type of misclassification has a serious meaning; that is, the CAD system does not signal the presence of a pathological pattern. The number of false negatives and false positives of GG and HC classes also decreased with the presence of DLac features increasing the correct classification of GG and HC instances. SD decreases in all metrics and classes, except for sensitivity of EMP class. So, the DLac features also improved the classifier stability.

The highest percentage of misclassified examples occurred among NOR and GG classes. Almost fifty percent (47.8%) of all the classification errors were due to incorrect classifications between these two classes. Figure 6 illustrates some random examples of normal ROIs that were classified as GG, on left column, and examples of ROIs with GG pattern that were classified as NOR, on right column. Although GG opacities are characterized by areas of increased attenuation, sometimes they are not dense enough to “hide” the broncho-vascular markings, especially in the initial phases of ILD diseases, associated with the presence of GG patterns.

4. Conclusions

A CAD scheme applied to HRCT images of the chest was proposed for the classification of healthy lung regions and with the presence of ILD. A texture analysis was performed to describe the lung patterns in study. Texture information of each ROI was represented by features extracted using a multiscale DLac approach combined with features obtained by classical statistical texture analysis methods. Feature selection and SVM training was performed over a 10-fold stratified

CV. The performance evaluation of the classifier model was assessed using an independent test set.

Experimental results showed that DLac features improve the performance of the proposed CAD system in both suggested scenarios: normal versus pathological and multiclass. In this case, the number of false negatives and false positives of NOR class decreased, as well as the misclassification between instances of pathological classes. Differentiating the normal pattern from pathological patterns, the classifier accuracy improved with an average of 1.4% when DLac features were considered, resulting in a correct classification of $95.8 \pm 2.2\%$ of all instances. In the multiclass scenario the overall accuracy was improved from $91.9 \pm 1.9\%$ to $94.5 \pm 2.1\%$ due to the presence of DLac features. The performance of the proposed CAD highlights the good discriminatory properties of extracted DLac features, making it suitable to integrate clinical applications for the classification of patterns associated with ILD.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

The authors thank Dr. Luísa Teixeira and Dr. Miguel Sêco from Radiology Department of Coimbra Hospital and University Centre, for their medical knowledge and assistance. This work is financed by the FCT-Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within Project UID/EEA/50014/2013.

References

- [1] E. A. Hoffman, J. M. Reinhardt, M. Sonka et al., "Characterization of the interstitial lung diseases via density-based and texture-based analysis of computed tomography images of lung structure and function," *Academic Radiology*, vol. 10, no. 10, pp. 1104–1118, 2003.
- [2] J. A. Verschakelen and W. D. Wever, *Computed Tomography of the Lung—A Pattern Approach*, Springer, Berlin, Germany, 2007.
- [3] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 385–405, 2006.
- [4] R. Uppaluri, E. A. Hoffman, M. Sonka, P. G. Hartley, G. W. Hunninghake, and G. McLennan, "Computer recognition of regional lung disease patterns," *American Journal of Respiratory and Critical Care Medicine*, vol. 160, no. 2, pp. 648–654, 1999.
- [5] A. Depeursinge, J. Iavindrasana, A. Hidki et al., "Comparative performance analysis of state-of-the-art classification algorithms applied to lung tissue categorization," *Journal of Digital Imaging*, vol. 23, no. 1, pp. 18–30, 2010.
- [6] W. Zhao, R. Xu, Y. Hirano, R. Tachibana, and S. Kido, "Classification of diffuse lung diseases patterns by a sparse representation based method on HRCT images," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5457–5460, IEEE, July 2013.
- [7] M. Anthimopoulos, S. Christodoulidis, A. Christe, and S. Mougiakakou, "Classification of interstitial lung disease patterns using local DCT features and random forest," in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '14)*, pp. 6040–6043, IEEE, Chicago, Ill, USA, August 2014.
- [8] M. Anthimopoulos, S. Christodoulidis, A. Christe, and S. Mougiakakou, "Computer-aided diagnosis of interstitial lung diseases based on computed tomography image analysis," in *Computational Optical Biomedical Spectroscopy and Imaging*, pp. 175–200, CRC Press, 2015.
- [9] 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," *Investigative Radiology*, vol. 32, no. 9, pp. 566–574, 1997.
- [10] S. Delorme, M. A. Keller-Reichenbecher, I. Zuna, W. Schlegel, and G. V. Kaick, "Interstitial lung disease a quantitative study using the adaptive multiple feature method," *American Journal of Respiratory and Critical Care Medicine*, vol. 159, no. 2, pp. 519–1073, 1999.
- [11] 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, 2003.
- [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," *Medical Physics*, vol. 30, no. 12, pp. 3081–3090, 2003.
- [13] Y. Xu, M. Sonka, G. McLennan, A. J. Guo, and E. A. Hoffmam, "Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM)," *IEEE Transactions on Medical Imaging*, vol. 16, no. 8, pp. 969–987, 2006.
- [14] A. Zaia, R. Eleonori, P. Maponi, R. Rossi, and R. Murri, "MR imaging and osteoporosis: fractal lacunarity analysis of trabecular bone," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 484–489, 2006.
- [15] A. Zaia, "Fractal lacunarity of trabecular bone and magnetic resonance imaging: new perspectives for osteoporotic fracture risk assessment," *World Journal of Orthopedics*, vol. 6, no. 2, pp. 221–235, 2015.
- [16] O. S. Al-Kadi and D. Watson, "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 7, pp. 1822–1830, 2008.
- [17] V. S. Charisis, L. J. Hadjileontiadis, and G. D. Sergiadis, "Enhanced ulcer recognition from capsule endoscopic images using texture analysis," in *New Advances in the Basic and Clinical Gastroenterology*, T. Brzozowski, Ed., pp. 185–210, InTech, Rijeka, Croatia, 2012.
- [18] V. Vasconcelos, L. Marques, J. S. Silva, and J. Barroso, "Lacunarity analysis of pulmonary emphysema in high-resolution CT images," in *Proceedings of the 8th Iberian Conference on Information Systems and Technologies (CISTI '13)*, pp. 1–5, Lisboa, Portugal, June 2013.
- [19] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [20] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [21] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1–2, pp. 169–186, 2003.

- [22] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, New York, NY, USA, 2nd edition, 2002.
- [23] R. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.
- [24] V. Vasconcelos, J. Silva, L. Marques, and J. Barroso, "Statistical textural features for classification of lung emphysema in CT images: a comparative study," in *Proceedings of the 5th Iberian Conference on Information Systems and Technologies*, pp. 1–5, Santiago de Compostela, Spain, June 2010.
- [25] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172–179, 1975.
- [26] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognition Letters*, vol. 11, no. 6, pp. 415–419, 1990.
- [27] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level—run length distributions," *Pattern Recognition Letters*, vol. 12, no. 8, pp. 497–502, 1991.
- [28] B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, New York, NY, USA, 1983.
- [29] B. B. Mandelbrot, "A fractal's lacunarity, and how it can be tuned and measured," in *Fractals in Biology and Medicine*, T. F. Nonnenmacher, G. A. Losa, and E. R. Weibel, Eds., Mathematics and Biosciences in Interaction, pp. 8–21, Birkhäuser, Boston, Mass, USA, 1994.
- [30] Y. Gefen, Y. Meir, B. B. Mandelbrot, and A. Aharony, "Geometric implementation of hypercubic lattices with noninteger dimensionality by use of low lacunarity fractal lattices," *Physical Review Letters*, vol. 50, no. 3, pp. 145–148, 1983.
- [31] R. E. Plotnick, R. H. Gardner, W. W. Hargrove, K. Prestegard, and M. Perlmutter, "Lacunarity analysis: a general technique for the analysis of spatial patterns," *Physical Review E*, vol. 53, no. 5, pp. 5461–5468, 1996.
- [32] P. Dong, "Test of a new lacunarity estimation method for image texture analysis," *International Journal of Remote Sensing*, vol. 21, no. 17, pp. 3369–3373, 2000.
- [33] C. Allain and M. Cloitre, "Characterizing the lacunarity of random and deterministic fractal sets," *Physical Review A*, vol. 44, no. 6, pp. 3552–3558, 1991.
- [34] N. Sarkar and B. B. Chaudhuri, "An efficient approach to estimate fractal dimension of textural images," *Pattern Recognition*, vol. 25, no. 9, pp. 1035–1041, 1992.
- [35] G. Dreyfus and I. Guyon, "Assessment methods," in *Feature Extraction: Foundations and Applications*, I. Guyon, M. Nikravesh, S. Gunn, and L. A. Zadeh, Eds., vol. 207 of *Studies in Fuzziness and Soft Computing: I*, pp. 65–86, Springer, Berlin, Germany, 2006.
- [36] N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley-Interscience, Hoboken, NJ, USA, 1998.
- [37] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [38] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [39] V. Vasconcelos, J. S. Silva, and J. Barroso, "CAD lung system: texture based classifier of pulmonary pathologies," in *Proceedings of the 8th Iberian Conference on Information Systems and Technologies*, pp. 383–386, June 2009.
- [40] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI '95)*, vol. 2, pp. 1137–1143, Montreal, Canada, August 1995.
- [41] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.
- [42] Y. Lee, J. B. Seo, J. G. Lee, S. S. Kim, N. Kim, and S. H. Kang, "Performance testing of several classifiers for differentiating obstructive lung diseases based on texture analysis at high-resolution computerized tomography (HRCT)," *Computer Methods and Programs in Biomedicine*, vol. 93, no. 2, pp. 206–215, 2009.
- [43] L. J. Hadjileontiadis, "A texture-based classification of crackles and squawks using lacunarity," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 718–732, 2009.
- [44] C.-C. Chang and C.-J. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.