

Automated 3D Interstitial Lung Disease Extent Quantification: Performance Evaluation and Correlation to PFTs

Alexandra Kazantzi · Lena Costaridou · Spyros Skiadopoulos · Panayiotis Korfiatis · Anna Karahaliou · Dimitris Daoussis · Andreas Andonopoulos · Christina Kalogeropoulou

Published online: 22 January 2014
© Society for Imaging Informatics in Medicine 2014

Abstract In this study, the performance of a recently proposed computer-aided diagnosis (CAD) scheme in detection and 3D quantification of reticular and ground glass pattern extent in chest computed tomography of interstitial lung disease (ILD) patients is evaluated. CAD scheme performance was evaluated on a dataset of 37 volumetric chest scans, considering five representative axial anatomical levels per scan. CAD scheme reliability analysis was performed by estimating agreement (intraclass correlation coefficient, ICC) of automatically derived ILD pattern extent to semi-quantitative disease extent assessment in terms of 29-point rating scale provided by two expert radiologists. Receiver operating characteristic (ROC) analysis was employed to assess CAD scheme accuracy in ILD pattern detection in terms of area under ROC curve (A_z). Correlation of reticular and ground glass volumetric pattern extent to pulmonary function tests (PFTs) was also investigated. CAD scheme reliability was substantial for ILD extent (ICC=0.809) and distinct reticular pattern extent (0.806) and moderate for distinct ground glass pattern extent (0.543), performing within inter-observer agreement. CAD scheme demonstrated high accuracy in detecting total ILD ($A_z=0.950\pm 0.018$), while accuracy in detecting distinct reticular and ground glass patterns was 0.920 ± 0.023 and 0.883 ± 0.024 ,

respectively. Moderate and statistically significant negative correlation was found between reticular volumetric pattern extent and diffusing capacity, forced expiratory volume in 1 s, forced vital capacity, and total lung capacity ($R=-0.581, -0.513, -0.494, \text{ and } -0.446$, respectively), similar to correlations found between radiologists' semi-quantitative ratings with PFTs. CAD-based quantification of disease extent is in agreement with radiologists' semi-quantitative assessment and correlates to specific PFTs, suggesting a potential imaging biomarker for ILD staging and management.

Keywords Interstitial lung disease · Disease extent assessment · Automated 3D disease extent quantification · Semi-quantitative scoring · Pulmonary function tests

Introduction

Chest computed tomography (CT) imaging is a powerful modality for detection, diagnosis, and follow-up of interstitial lung disease (ILD) [1]. Monitoring of ILD progression is based on the estimation of disease extent in CT image data and by means of pulmonary function tests (PFTs) [1,2]. Disease extent in chest CT of patients with collagen vascular disease (CVD) constitutes a strong predictor of mortality [3–5], while the accuracy of disease extent assessment affects the identification of change in follow-up and treatment decision making [4,5].

Semi-quantitative approaches of ILD extent estimation are based on visual assessment of ILD patterns in high-resolution CT or volumetric data and subsequent scoring by means of ordinal rating scales [6–9]. Likert scale accounts for a four-point or five-point rating scale of disease extent (e.g., 0–25, 26–50, 51–75, 76–100 %) [6]. Desai et al. [7] proposed a 21-point rating scale (0 up to

A. Kazantzi · C. Kalogeropoulou (✉)
Department of Radiology, School of Medicine, University of Patras,
26504 Patras, Greece
e-mail: rat@med.upatras.gr

L. Costaridou · S. Skiadopoulos · P. Korfiatis · A. Karahaliou
Department of Medical Physics, School of Medicine, University of
Patras, 26504 Patras, Greece

D. Daoussis · A. Andonopoulos
Division of Rheumatology, Department of Internal Medicine,
School of Medicine, University of Patras, 26504 Patras, Greece

100 %, with step 5 %) reflecting disease extent in three or five representative lung anatomical levels. However, semi-quantitative approaches of ILD extent estimation demonstrate moderate inter- and intra-observer agreement, and they do not provide information for disease localization [10–14].

Automated image analysis tools, also referred as computer-aided diagnosis (CAD) schemes, focused on the quantification of ILD extent have been proposed [15–22] to improve efficiency, accuracy, and reproducibility in disease extent assessment. Furthermore, the contribution of CAD schemes in facilitating disease assessment in case of volumetric chest datasets originating from multi-detector CT (MDCT) has been recently highlighted [23–25].

CAD schemes proposed for ILD extent quantification are based on gray level thresholding techniques [15–17] and texture-based voxel/pixel classification schemes of lung parenchyma [18–22]. Texture-based image analysis approaches have also been exploited for the differentiation of automatically [26,27] or manually [28–30] defined lung parenchyma regions/voxels of interest (VOIs) into various ILD patterns, allowing for a coarse identification of disease patterns and not for a detailed disease extent quantification scheme.

CAD schemes for ILD extent quantification are differentiated with respect to the 2D or 3D disease extent output. Most CAD schemes exploit high-resolution CT (HRCT) image acquisition protocols [18,19,21,22] which limit the coverage of the entire lung parenchyma volume and lack anatomic comparability in case of follow-up studies [25,26], while only a few CAD schemes are applicable on MDCT image data thus providing volumetric quantification of disease extent [15–17,20].

Automated ILD extent quantification studies are also differentiated with respect to the pattern being quantified. Reticular and ground glass account for the ILD patterns that mainly reflect disease progression. The importance of estimating the contribution of each specific pattern in disease progression has been recently highlighted [9], rendering the development of CAD schemes capable of identifying and quantifying distinct reticular and ground glass patterns crucial. Nevertheless, only a few CAD schemes allow for the quantification of distinct reticular and ground glass patterns [18,20,22]. The development of CAD schemes that provide 3D quantification of ILD extent and are further capable of quantifying distinct reticular and ground glass patterns accounts for an ongoing research issue [20].

In the framework of establishing volumetric disease extent as an imaging biomarker for ILD progression and management, it is important to investigate the correlation of

automatically derived (CAD-based) disease extent with PFTs also reflecting disease progression [15–17,19,21]. Correlation of CAD-based disease extent with PFTs has also been investigated in chronic obstructive pulmonary disease patients, in terms of emphysema pattern [28,31]. However, investigating the correlation of distinct reticular and ground glass volumetric pattern extent to PFTs is currently under-researched.

The aim of the current study is to evaluate the performance of a recently proposed CAD scheme in detection and 3D quantification of reticular and ground glass pattern extent in MDCT datasets. Performance evaluation considers CAD scheme reliability analysis, by estimating agreement of automatically derived ILD pattern extent to semi-quantitative assessment and ROC analysis to assess system accuracy in ILD pattern detection. Correlation of reticular and ground glass volumetric pattern extent to PFTs is also investigated.

Materials and Methods

Clinical Dataset

Forty-six patients with diagnosis of CVD, meeting the criteria of the American College of Rheumatology [32], were referred to our institution for chest CT scan. Ethics committee approval was obtained before beginning of the study. Thirty-seven scans were retained from the initial case sample of 46 patient scans, exclusively exhibiting imaging findings of ground glass and reticular patterns, including honeycombing. The excluded nine scans demonstrated consolidation, nodular, and emphysematous patterns. The sample analyzed consisted of 37 patients (27 females and ten males) with mean age of 57.3 years and mean disease duration 7.6 years. Fifteen patients were diagnosed with scleroderma, ten patients with rheumatoid arthritis, two with systemic lupus erythematosus, one with polymyositis, one with Sjogren syndrome, and eight mixed connective tissue disease. Informed consent was obtained from all subjects participating in this study. For each patient scan, five representative lung anatomic levels were selected to assess disease extent: (a) origin of great vessel, (b) carina, (c) pulmonary venous confluence, (d) between pulmonary venous confluence and 1 cm above the right hemidiaphragm, and (e) 1 cm above the right hemidiaphragm. In total, a case sample of 185 axial slices (5 slices×37 patients) was generated.

Image Acquisition

All patients were scanned using tube voltage of 120 kVp, rotation time of 0.5 s, automatic modulation of mA,

collimation thickness of 16×0.625 mm, and slice thickness of 1.25 mm, obtaining volumetric data at full inspiration in supine position, in a 16-row MDCT scanner (GE Lightspeed 16, General Electric Medical Systems, Milwaukee, WI, USA). Each scan volume comprised of approximately 200–250 slices per patient. The mean volume CT dose index and the mean dose–length product were 11.0 mGy and 268.3 mGy cm, respectively. Assuming 0.017 mSv/mGy cm for a standard chest CT examination, the effective radiation dose for the volumetric chest CT protocol used was 4.6 mSv, complying with European Working Group for Guidelines on Quality Criteria in CT [33].

Semi-quantitative ILD Extent Assessment

Two radiologists with 20 and 10 years of experience in chest CT (C.K. and A.K., respectively), who were aware of patient history, provided independently semi-quantitative disease extent assessments (RAD_1 and RAD_2) for reticular pattern, ground glass pattern, and total disease extent for each of the 185 axial slices of the dataset (originating from 37 patient CT scans). In order to assess intra-observer agreement (reproducibility), the same interpretation scenario was repeated after a time interval of 1 month by the first radiologist (RAD'_1).

Semi-quantitative disease extent assessment was also provided by the same two radiologists in consensus (RAD_{cons}) after a time interval of 2 months. To achieve consensus, the two radiologists discussed and agreed on pattern type and extent per slice.

Semi-quantitative disease extent assessment was performed by means of a 29-point rating scale, derived by modifying the 21-point rating scale proposed by Desai et al. [7]. Specifically, for each axial slice of patient scan (five slices have been considered per patient scan, each one corresponding to one representative anatomical lung level, as previously defined), disease extent was provided as a percentage of the lung parenchyma area, ranging from 0 to 10 % in 1 % step and ranging from 10 to 100 % in 5 % step. This modification allows for a more detailed rating in case of limited disease extent (0 up to 10 % in 1 % step), as opposed to the coarse rating provided by Desai et al. [7] in the same interval (0 up to 10 % in 5 % step). The time required for disease extent estimation of each axial slice was approximately 2–3 min for each radiologist.

CAD-Based 3D ILD Extent Quantification

Disease extent was provided by a recently proposed computerized ILD quantification scheme [20]. Following, an overview of the system methodology is provided. A two-

stage architecture is adopted, comprised of (a) pre-processing and (b) texture-based voxel classification of the lung parenchyma volume.

The pre-processing stage involves 3D lung field segmentation [34,35] and 3D vessel tree segmentation [36] in order to isolate the lung parenchyma volume subsequently subjected to voxel classification. Specifically, lung-field segmentation is achieved by 3D automated gray-level thresholding combined with an edge-highlighting wavelet preprocessing step [34], followed by a texture-based border refinement step [35]. The vessel tree volume is identified by a combined scheme based on 3D multi-scale filtering [36] and subsequently removed from lung field volume, resulting in the lung parenchyma volume.

In the second stage, the lung parenchyma volume is sampled by a sliding overlapping VOI ($21 \times 21 \times 21$ pixels), whose center voxel is assigned into one of three classes (normal, reticular, ground glass) according to local texture properties, employing a k -nearest neighbor classifier ($k=10$). Local texture properties are captured by ten 3D gray-level co-occurrence texture features selected from 130 initially tested features (mean and range of 13 co-occurrence feature values over 13 orientations, considering five displacement values), by means of stepwise discriminant analysis. The training set of the voxel-classification scheme comprised of 350 cubic VOIs (VOI, $21 \times 21 \times 21$ pixels), representing patterns corresponding to reticular (150) and ground glass (100) patterns, as well as normal lung parenchyma (100). The 350 cubic VOIs were defined by expert radiologists, from lung CT data of five patients presenting with ILD and three normal ones, of a patient cohort distinct from the 37 patients considered in the current study.

The CAD scheme allows for the visualization of abnormal lung parenchyma areas and normal tissue in 3D volumes or 2D slices by means of color coding of reticular and ground glass patterns. Quantification of disease extent is provided in terms of percentage of reticular pattern extent, ground glass pattern extent, and total disease extent (considering both reticular and ground glass patterns) in 3D volumes or 2D slices.

Specifically, disease extent is expressed as percentage with respect to segmented lung parenchyma volume according to Eqs. (1)–(3).

$$\text{Reticular Extent} = \frac{\text{No. of Reticular Pattern Voxels}}{\text{No. of Lung Parenchyma Voxels}} \quad (1)$$

$$\text{Ground Glass Extent} = \frac{\text{No. of Ground Glass Pattern Voxels}}{\text{No. of Lung Parenchyma Voxels}} \quad (2)$$

$$\text{Total Disease Extent} = \frac{\text{No. of Ground Glass + Reticular Pattern Voxels}}{\text{No. of Lung Parenchyma Voxels}} \quad (3)$$

Disease extent is also provided on a slice basis, as percentage with respect to segmented lung parenchyma pixels in a single slice. In this case, Eqs. 1–3 stand by replacing the term “voxels” by the term “pixels.”

Pulmonary Function Tests

PFTs account for a set of indexes reflecting the lung functional capacity, commonly measured through spirometry. PFT indexes are expressed as a ratio of the corresponding predicted value based on the patients’ height, age, sex, and ethnic. Lung function is defined as “normal” when PFT index values are >85 % of the corresponding predicted values. ILD results in restrictive changes in lung function with reduced respiratory volumes, correlating with decrease of PFTs [37].

In the current study, pulmonary function testing was performed according to the European Respiratory Society and the American Thoracic Society criteria [37,38] with the patient at rest in a sitting position (Vmax 22, SensorMedics, Yorba Linda, CA, USA; PFDX, MedGraphics, St. Paul, MN, USA). PFT data were obtained within 60 days (mean=12±11 days; range=[−21, 56]days) from the day of CT image data acquisition.

The following indexes were obtained: forced vital capacity (FVC), forced expiratory volume in 1 s (FEV₁), total lung capacity (TLC), and diffusing capacity (DL_{CO}). The DL_{CO} index was assessed according to single-breath carbon monoxide uptake.

Statistical Analysis

CAD Scheme Reliability

Reliability analysis [39] was utilized in order to assess the agreement between CAD scheme and semi-quantitative disease assessment by the two radiologists in consensus (CAD scheme vs. RAD_{cons}), as well as between CAD scheme and semi-quantitative disease assessment by each radiologist independently (CAD scheme vs. RAD₁, CAD scheme vs. RAD₂), in quantification of total, reticular, and ground glass disease extent. Inter- and intra-observer reproducibility in semi-quantitative disease assessment (RAD₁ vs. RAD₂, RAD₁ vs. RAD₁'), as well as the agreement in semi-quantitative disease assessment between radiologists’ ratings in consensus and radiologists’ independent ratings (RAD_{cons} vs. RAD₁, RAD_{cons} vs. RAD₂), was also estimated by means of reliability analysis.

In reliability analysis, intraclass correlation coefficient (ICC) and its corresponding 95 % confidence intervals (CI) were calculated for total, reticular, and ground glass disease. The degree of agreement was scaled as almost perfect (ICC=[0.81–1.00]), substantial (ICC=[0.61–0.81]), moderate (ICC=[0.41–0.61]), or weak (ICC=[0.21–0.41]) [39].

CAD Scheme Detection Accuracy

CAD scheme accuracy in detection of total disease, as well as distinct reticular and ground glass patterns, was evaluated in terms of receiver operating characteristic (ROC) analysis [40]. Considering total disease detection accuracy evaluation, a ROC curve was generated by considering each automatically derived total disease extent value as cutoff point (decision threshold) and subsequently calculating the sensitivity and specificity associated with each cutoff point. Slices with automatically derived total disease extent values greater than or equal to the cutoff point are considered “abnormal” (positive), and slices with automatically derived total disease extent values less than the cutoff point are considered “normal” (negative). If a slice is abnormal according to ground truth and is characterized as “abnormal,” then it is counted as true positive. If a slice is characterized as “normal” while being abnormal according to ground truth, it is counted as a false negative. If the slice is negative according to ground truth and is characterized as “normal,” it is counted as a true negative; if it is characterized as “abnormal,” then it is counted as a false positive. The visual ratings of the two radiologists in consensus (RAD_{cons}), as well as ratings of the two radiologists considered independently (RAD₁, RAD₂), were used as “ground truth.” Specifically, the 185 slices were characterized with respect to total disease extent as normal (i.e., visual rating extent=0 %, 36 slices) or abnormal (i.e., visual scoring extent ≥1 %, 149 slices) [19].

The same approach was utilized for evaluating CAD scheme detection accuracy with respect to distinct reticular and ground glass patterns. Specifically, for evaluating CAD scheme reticular pattern detection accuracy, a ROC curve was generated by considering each automatically derived reticular pattern disease extent value as cutoff point (decision threshold) and subsequently calculating the sensitivity and specificity associated with each cutoff point. In case of reticular pattern, “ground truth” considered 53 slices characterized as normal (i.e., visual rating of reticular extent 0 %) and 132 slices characterized as abnormal (i.e., visual rating of reticular extent ≥1 %).

For evaluating CAD scheme ground glass pattern detection accuracy, a ROC curve was generated by considering each automatically derived ground glass pattern disease extent value as cutoff point (decision threshold) and subsequently calculating the sensitivity and specificity associated with each cutoff point. In case of ground glass pattern, “ground truth” considered 72 slices characterized as normal (i.e., visual rating of ground glass extent 0 %) and 113 slices characterized as abnormal (i.e., visual rating of ground glass extent ≥ 1 %).

The area under the empirical ROC curve (A_z), standard error (SE), as well as the asymmetric 95 % CI values were calculated for total disease, reticular, and ground glass patterns. Both CAD scheme reliability and detection accuracy analyses considered quantification of 2D extent of the disease, allowing for a detailed performance evaluation approach on a slice basis (185 slices in total were considered).

Correlation to PFTs

Pearson correlation coefficient (R) was employed to determine correlation between automatically derived (CAD-based) disease extent and PFT indexes. Specifically, correlation of 3D (volumetric) extent of total disease, of reticular, and of ground glass patterns (Eqs. 1–3) with FVC, FEV₁, TLC, and DL_{CO} was investigated.

Pearson correlation coefficient (R) was also employed to determine the correlation between radiologists’ ratings in consensus (RAD_{cons}) for total disease, reticular and ground glass pattern extent, and PFT indexes. For each patient scan, semi-quantitative assessment of disease volumetric extent was derived by averaging radiologists’ ratings reflecting disease extent for each of the five representative axial slices [7]).

Correlation was characterized as very strong [−1.0, −0.9], strong (−0.9, −0.7], moderate (−0.7, −0.4], weak (−0.4, −0.2], or very weak to negligible (−0.2, −0.0]. A P value less than 0.05 was accepted to indicate statistical significance. Statistical analysis was performed using the IBM SPSS Statistics software package (SPSS Release 20.0, SPSS Inc., Chicago, IL, USA).

Results

Radiologist Reproducibility

Table 1 summarizes results of the reliability analysis performed to assess inter- and intra-observer agreement in semi-quantitative extent assessment of total disease, reticular, and ground glass patterns. Agreement in semi-quantitative disease assessment between radiologists’ ratings in consensus and radiologists’ independent ratings (RAD_{cons} vs. RAD₁, RAD_{cons} vs. RAD₂) is also provided.

Inter-observer agreement of radiologists was substantial for total disease (ICC=0.776, CI=[0.674–0.843]) and reticular pattern (ICC=0.727, CI=[0.603–0.808]), while moderate for ground glass pattern (ICC=0.472, CI=[0.352–0.577]). Intra-observer agreement of radiologists was substantial for total disease (ICC=0.785, CI=[0.711–0.840]) and reticular pattern (ICC=0.736, CI=[0.651–0.800]), while moderate for ground glass pattern (ICC=0.538, CI=[0.427–0.633]).

Agreement between radiologists’ ratings in consensus and radiologists’ independent ratings (RAD_{cons} vs. RAD₁, RAD_{cons} vs. RAD₂) was almost perfect for total disease extent assessment (ICC=0.839, CI=[0.790–0.877] and ICC=0.852, CI=[0.788–0.895], respectively) and for reticular pattern in case of RAD₁ (ICC=0.853, CI=[0.808–0.888]), while being substantial for reticular pattern in case of RAD₂ (ICC=0.789, CI=[0.695–0.851]) and for ground glass pattern (ICC=0.674, CI=[0.588–0.746] and ICC=0.610, CI=[0.510–0.692], respectively).

CAD Scheme Reliability

Figure 1 provides application examples of the computerized ILD quantification scheme on MDCT scans of two patients with scleroderma disease. Visualization of abnormal lung parenchyma areas and normal tissue is provided in 3D representation and 2D axial slices, coded as green overlay for ground glass pattern and blue overlay for reticular pattern.

Table 1 Inter- and intra-observer agreement in semi-quantitative disease assessment, in terms of ICC and corresponding 95 % CI, between radiologists’ ratings (RAD₁ vs. RAD₂) as well as between radiologist first and second interpretation (RAD₁ vs. RAD’₁), for total disease,

reticular, and ground glass pattern extent. Agreement in semi-quantitative disease assessment between radiologists’ ratings in consensus and radiologists’ independent ratings (RAD_{cons} vs. RAD₁, RAD_{cons} vs. RAD₂) is also provided

| Extent | Inter-observer | | Intra-observer | | RAD _{cons} vs. RAD ₁ | | RAD _{cons} vs. RAD ₂ | |
|--------------|----------------|-------------|----------------|-------------|--|-------------|--|-------------|
| | ICC | CI | ICC | CI | ICC | CI | ICC | CI |
| Total | 0.776 | 0.674–0.843 | 0.785 | 0.711–0.840 | 0.839 | 0.790–0.877 | 0.852 | 0.788–0.895 |
| Reticular | 0.727 | 0.603–0.808 | 0.736 | 0.651–0.800 | 0.853 | 0.808–0.888 | 0.789 | 0.695–0.851 |
| Ground glass | 0.472 | 0.352–0.577 | 0.538 | 0.427–0.633 | 0.674 | 0.588–0.746 | 0.610 | 0.510–0.692 |

ICC intraclass correlation coefficient, CI confidence intervals

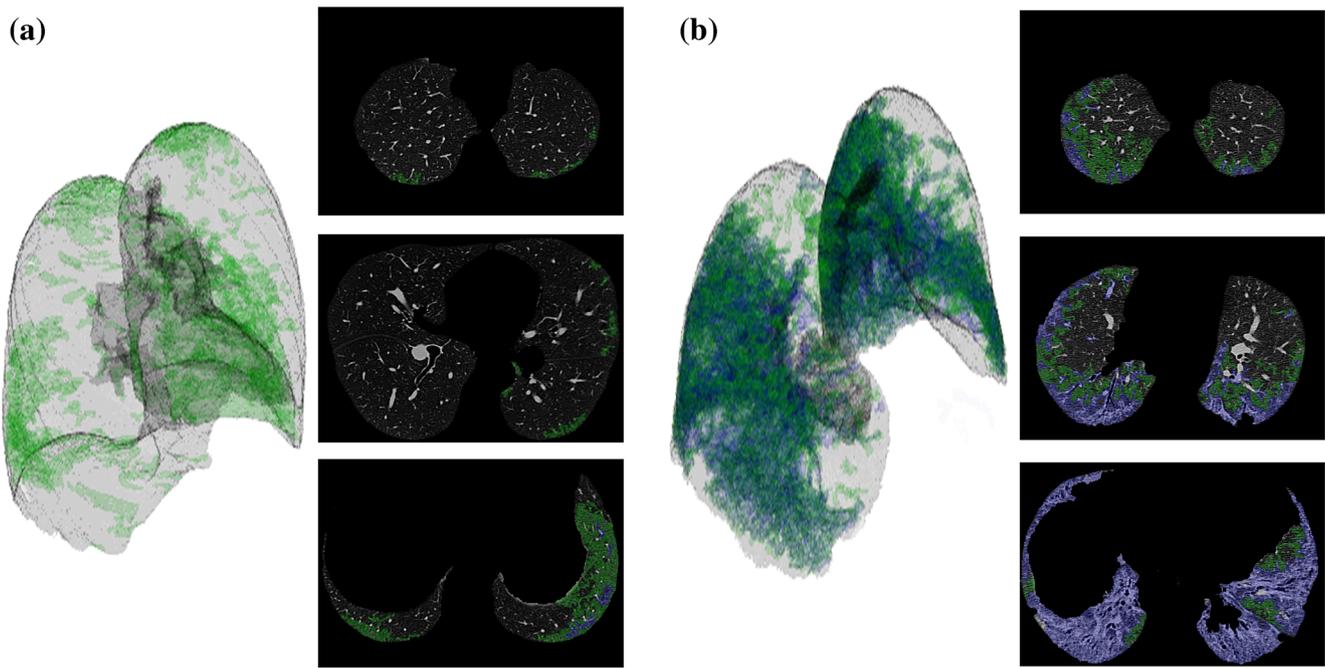


Fig. 1 Application examples of the computerized ILD quantification scheme on two MDCT patient scans. Visualization of abnormal lung parenchyma and normal tissue in 3D and 2D representation (three representative axial anatomic levels), as provided by the computerized ILD quantification scheme. *Green overlay* corresponds to ground glass pattern, and *blue overlay* corresponds to reticular pattern. *Gray overlay* in

3D representation corresponds to normal lung parenchyma. **a** A 48-year-old female patient with scleroderma disease presenting with mild ILD with predominant pattern ground glass over reticular. **b** A 39-year-old male patient with 15 years of scleroderma presenting with extensive ILD with mixed reticular and ground glass pattern and severe lung impairment

Gray overlay in 3D representation corresponds to normal lung parenchyma.

Table 2 summarizes results of reliability analysis, in terms of ICC and corresponding CI, reflecting agreement between CAD scheme and semi-quantitative disease assessment by each radiologist independently (CAD scheme vs. RAD₁, CAD scheme vs. RAD₂), as well as between CAD scheme and semi-quantitative disease assessment provided by two radiologists' in consensus (CAD scheme vs. RAD_{cons}), for total disease, reticular, and ground glass pattern extent.

When radiologists' ratings are considered independently, CAD scheme reliability is substantial with respect to total disease extent assessment (ICC=0.736, CI=[0.647–0.803] for RAD₁; ICC=0.653, CI=[0.336–0.801] for RAD₂). CAD

scheme reliability is substantial with respect to reticular pattern extent assessment (ICC=0.690, CI=[0.605–0.659] for RAD₁; ICC=0.615, CI=[0.393–0.738] for RAD₂) and weak for ground glass pattern extent assessment (ICC=0.354, CI=[0.222–0.473] for RAD₁; ICC=0.233, CI=[0.095–0.362] for RAD₂). In case of radiologists' ratings in consensus, CAD scheme reliability is substantial for total disease extent (ICC=0.809, CI=[0.599, 0.894]) and reticular pattern extent (ICC=0.806, CI=[0.714, 0.865]) and moderate for ground glass pattern extent assessment (ICC=0.543, CI=[0.405, 0.652]).

Furthermore, CAD scheme agreement with consensus radiologists' ratings for total disease, reticular pattern, and ground glass pattern extent assessment ranges within inter-

Table 2 CAD scheme reliability, in terms of ICC and corresponding 95 % CI, between CAD scheme and radiologists' ratings with respect to total disease, reticular, and ground glass pattern extent

| Extent | CAD scheme vs. RAD _{cons} | | CAD scheme vs. RAD ₁ | | CAD scheme vs. RAD ₂ | |
|--------------|------------------------------------|-------------|---------------------------------|-------------|---------------------------------|-------------|
| | ICC | CI | ICC | CI | ICC | CI |
| Total | 0.809 | 0.599–0.894 | 0.736 | 0.647–0.803 | 0.653 | 0.336–0.801 |
| Reticular | 0.806 | 0.714–0.865 | 0.690 | 0.605–0.759 | 0.615 | 0.393–0.738 |
| Ground glass | 0.543 | 0.405–0.652 | 0.354 | 0.222–0.473 | 0.233 | 0.095–0.362 |

ICC intraclass correlation coefficient, CI confidence intervals

Table 3 CAD scheme detection accuracy in terms of A_z and SE, as well as asymmetric 95 % CI, with respect to total disease, reticular, and ground glass patterns, considering as “ground truth” radiologists’ ratings in consensus (RAD_{cons}) and as independent ratings (RAD₁, RAD₂)

| Extent | RAD _{cons} | | RAD ₁ | | RAD ₂ | |
|--------------|---------------------|-------------|------------------|-------------|------------------|-------------|
| | $A_z \pm SE$ | CI | $A_z \pm SE$ | CI | $A_z \pm SE$ | CI |
| Total | 0.950±0.018 | 0.900–0.976 | 0.925±0.020 | 0.874–0.956 | 0.931±0.024 | 0.866–0.965 |
| Reticular | 0.920±0.023 | 0.860–0.955 | 0.920±0.020 | 0.869–0.951 | 0.902±0.022 | 0.850–0.937 |
| Ground glass | 0.883±0.024 | 0.826–0.922 | 0.792±0.036 | 0.712–0.852 | 0.766±0.037 | 0.682–0.830 |

A_z area under empirical ROC curve, SE standard error, CI confidence intervals

observer agreement with semi-quantitative disease assessment (Table 1). CAD scheme agreement with independent radiologists’ ratings ranges within inter-observer agreement in the case of reticular pattern extent assessment.

CAD Scheme Detection Accuracy

Table 3 provides CAD scheme detection accuracy, in terms of A_z index, with respect to total disease, reticular, and ground glass patterns. Considering the two radiologists’ ratings in consensus (RAD_{cons}) as “ground truth,” the CAD scheme demonstrated high accuracy in detecting total disease ($A_z \pm SE = 0.950 \pm 0.018$). CAD scheme demonstrated higher accuracy in case of reticular pattern detection ($A_z \pm SE = 0.920 \pm 0.023$) as compared to ground glass pattern detection ($A_z \pm SE = 0.883 \pm 0.024$). The same trend was also observed when radiologists’ ratings are considered independently (RAD₁, RAD₂) as ground truth.

Figure 2 illustrates ROC curves corresponding to CAD scheme detection accuracy considering the two radiologists’

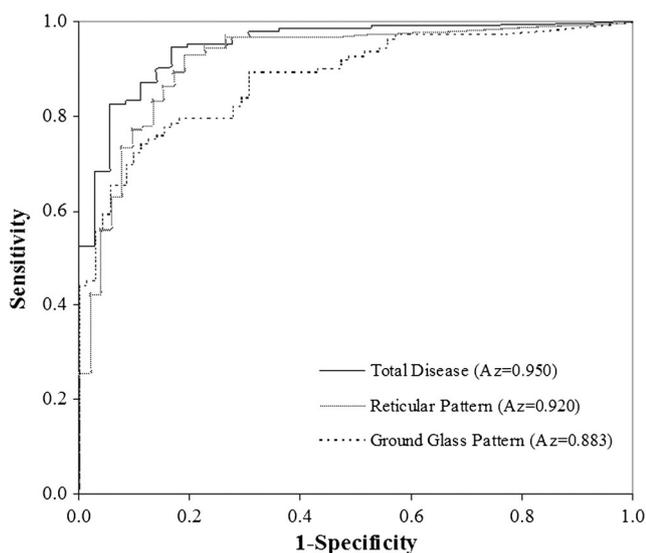


Fig. 2 Empirical ROC curves reflecting CAD scheme accuracy in the detection of total disease, reticular, and ground glass patterns, considering the two radiologists in consensus as “ground truth”

ratings in consensus (RAD_{cons}) as “ground truth.” The CAD scheme achieves sensitivity of 94.6 % and corresponding specificity of 83.3 % (overall accuracy 92.4 %) in detecting total disease, using the 0.984 % disease extent as cutoff point (threshold). The CAD scheme achieves sensitivity of 97.0 % and corresponding specificity of 73.6 % (overall accuracy 90.3 %) in detecting reticular patterns, using the 0.289 % disease extent cutoff point (threshold). The CAD scheme achieves sensitivity of 89.4 % and corresponding specificity of 69.4 % (overall accuracy 81.6 %) in detecting ground glass patterns, using the 0.551 % disease extent cutoff point (threshold).

Correlation Between CAD-Based ILD Extent Quantification and PFTs

Table 4 presents Pearson correlation (R) values reflecting correlation between automatically quantified 3D extent of total disease, reticular, and ground glass pattern with four PFT indexes (DL_{CO}, FEV₁, FVC, TLC), also being schematically illustrated in Fig. 3. CAD-based 3D extent of total disease demonstrated weak and statistically significant negative correlation with DL_{CO} ($R = -0.335$, $P = 0.043$). Correlation with respect to FEV₁, FVC, and TLC was weak to very weak without demonstrating statistical significance. CAD-based 3D extent of reticular pattern demonstrated moderate and

Table 4 Pearson correlation (R) and corresponding P values between CAD-based disease extent quantification and pulmonary function tests

| PFT | Total | | Reticular | | Ground Glass | |
|------------------|--------|--------------------|-----------|---------------------|--------------|-------|
| | R | P | R | P | R | P |
| DL _{CO} | -0.335 | 0.043 ^a | -0.581 | <0.001 ^a | 0.058 | 0.733 |
| FEV ₁ | -0.321 | 0.053 | -0.513 | 0.001 ^a | 0.006 | 0.973 |
| FVC | -0.263 | 0.116 | -0.494 | 0.002 ^a | 0.090 | 0.595 |
| TLC | -0.167 | 0.322 | -0.446 | 0.006 ^a | 0.210 | 0.213 |

PFT pulmonary function tests, DL_{CO} diffusing capacity, FEV₁ forced expiratory volume in 1 s, FVC forced vital capacity, TLC total lung capacity

^a Statistically significant

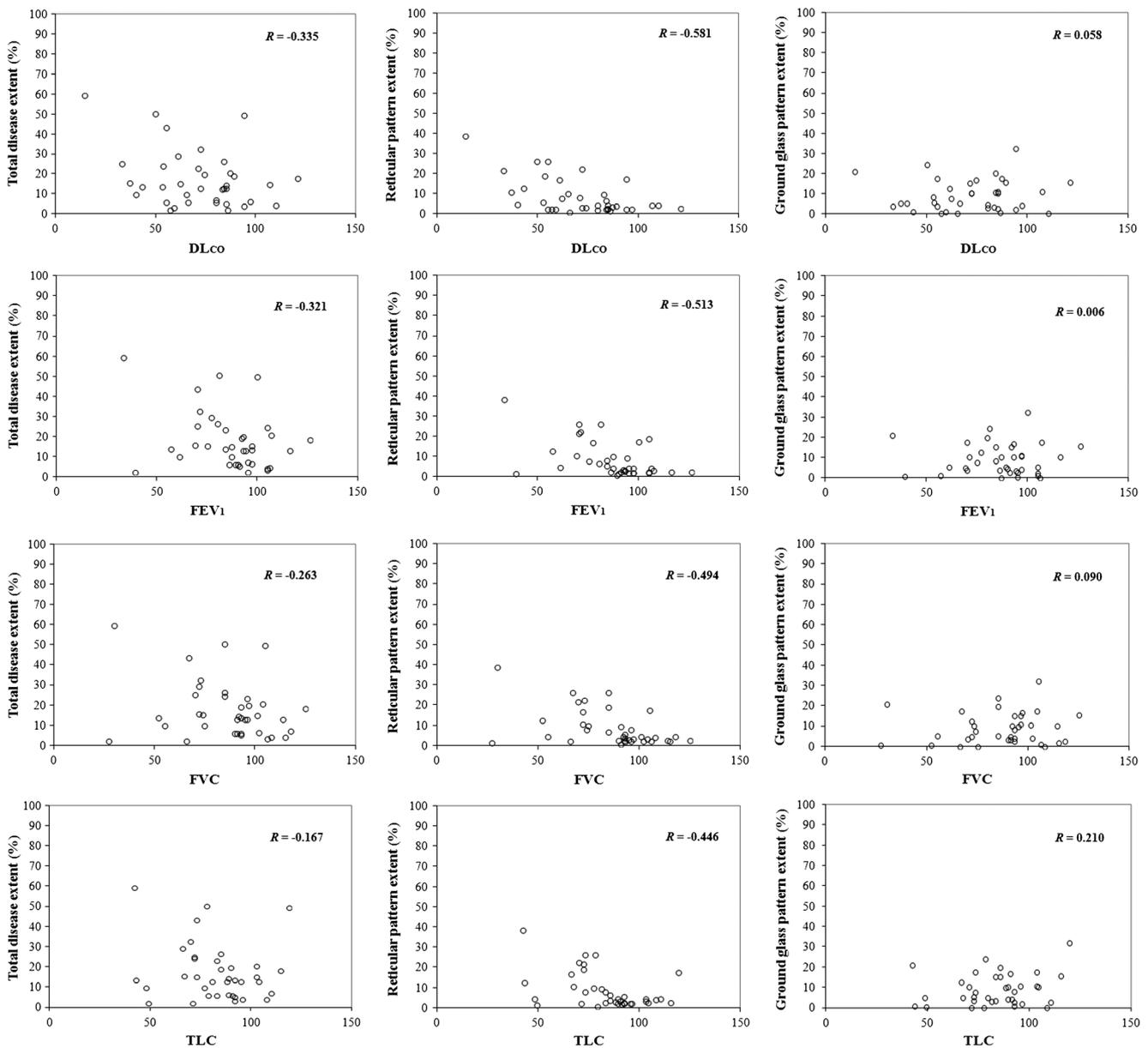


Fig. 3 Correlation between automatically quantified 3D extent of total disease, reticular, and ground glass pattern with PFT indexes (FVC, FEV₁, TLC, and DL_{CO})

statistically significant negative correlations with all PFT indexes studied (ranging from $R = -0.581$, $P < 0.001$ for DL_{CO} to $R = -0.446$, $P = 0.006$ for TLC). No statistically significant correlation was found between CAD-based 3D extent of ground glass pattern and PFT indexes studied.

Correlation Between Semi-quantitative ILD Extent Assessment and PFTs

Table 5 presents Pearson correlation (R) values reflecting correlation between semi-quantitative assessment of total, reticular, and ground glass disease extent as provided by in consensus radiologists' ratings, with

four PFT indexes (DL_{CO}, FEV₁, FVC, TLC). Schematic illustration of investigated correlation is provided in Fig. 4.

Total disease extent assessed semi-quantitatively demonstrated weak and statistically significant negative correlation with DL_{CO} ($R = -0.398$, $P = 0.015$). Correlation with respect to FEV₁, FVC, and TLC was weak without demonstrating statistical significance.

Reticular pattern extent assessed semi-quantitatively demonstrated weak to moderate and statistically significant negative correlations with all PFT indexes studied. Specifically, moderate correlation was observed with respect to DL_{CO} ($R = -0.485$, $P = 0.002$) and FEV₁ ($R =$

Table 5 Pearson correlation (R) and corresponding P values between semi-quantitative radiologists' ratings of disease extent provided by the two radiologists in consensus and pulmonary function tests

| PFT | Total | | Reticular | | Ground glass | |
|-----------|--------|--------------------|-----------|--------------------|--------------|-------|
| | R | P | R | P | R | P |
| DL_{CO} | -0.398 | 0.015 ^a | -0.485 | 0.002 ^a | -0.059 | 0.730 |
| FEV_1 | -0.303 | 0.068 | -0.411 | 0.012 ^a | 0.012 | 0.945 |
| FVC | -0.288 | 0.084 | -0.360 | 0.029 ^a | -0.031 | 0.856 |
| TLC | -0.204 | 0.225 | -0.326 | 0.049 ^a | 0.076 | 0.655 |

PFT pulmonary function tests, DL_{CO} diffusing capacity, FEV_1 forced expiratory volume in 1 s, FVC forced vital capacity, TLC total lung capacity

^a Statistically significant

-0.411, $P=0.012$), while weak with respect to FVC ($R=-0.360$, $P=0.029$) and TLC ($R=-0.326$, $P=0.049$). No statistically significant correlation was found between ground glass pattern extent assessed semi-quantitatively and PFT indexes studied.

Discussion

Estimation of ILD extent through MDCT imaging is fundamental for efficient disease management. Taking into account the subjectivity induced in semi-quantitative assessment of ILD extent, automated ILD quantification schemes contribute in objective and accurate volumetric disease extent estimation. Nevertheless, the identification of distinct ILD patterns by such automated quantification tools, as well as their correlation to well-established clinical factors of disease progression such as PFTs, remains an open research issue.

The current study focused on evaluating the performance of a texture-based lung parenchyma voxel classification scheme [20] in the detection and volumetric quantification of distinct reticular and ground glass pattern extent in 37 MDCT patient scans. The correlation of the quantified ILD patterns with four PFT indexes is also encountered.

Performance evaluation considered CAD scheme reliability analysis, by estimating agreement of automatically derived ILD pattern extent to semi-quantitative assessment. Results demonstrated that automatically derived total disease extent agrees substantially to semi-quantitative rating of radiologists (Table 2), also performing within intra- and inter-observer variability range (Table 1). Heterogeneity of the datasets analyzed renders direct comparison to previously reported studies not feasible. Furthermore, taking into account that most of the previously reported studies [15–17,19,21] have adopted the Pearson or Spearman (R) correlation coefficient to investigate agreement, as opposed to ICC adopted herein, the following

comparisons are only indicative of existing trends. Specifically, results reported herein are in accordance to Marten et al. [15,16] also demonstrating substantial agreement ($R=0.716$) of computer-derived total disease extent (as provided by a threshold-based method) to semi-quantitative scoring of 52 volumetric patient scans, reduced, however, as compared to inter-observer agreement ($R=0.89$). Shin et al. [17] reported moderate agreement ($R=0.53$) of total disease extent quantified by a threshold-based approach, with semi-quantitative assessment of disease extent in a total of 157 volumetric patient scans. Kim et al. [19] employing a texture-based approach for disease extent quantification reported moderate agreement ($R=0.60$) with semi-quantitative assessment of disease extent by means of a five-point rating scale in case sample of 129 HRCT patient scans. Rosas et al. [21] employed a texture-based approach for quantifying 2D disease extent in 86 HRCT patient scans and reported agreement of 0.562 (in terms of kappa coefficient) with radiologists' semi-quantitative assessment of total disease extent by means of five-point rating scale.

Results of CAD scheme reliability analysis also demonstrated that automatically derived disease extent of reticular pattern agrees substantially to semi-quantitative rating of radiologists, performing within intra- and inter-observer variability range (Table 1). CAD-derived disease extent of ground glass pattern demonstrated reduced agreement to semi-quantitative rating of radiologists, a task also characterized by increased intra- and inter-observer variability. This fact is probably attributed to the specific underlying pathology manifested by each ILD pattern. Reticular pattern corresponds to irreversible lung fibrosis. On the other hand, the pathology of ground glass is ambiguous: In some cases, it reflects inflammatory process thus reversible lung damage, while in other cases especially when combined with reticular, it reflects irreversible disease. These results are in accordance to Marten et al. [15], also considering distinct ILD patterns when investigating agreement of automatically derived disease extent to semi-quantitative assessment. Specifically, they demonstrated lack of agreement of CAD-derived ground glass pattern extent to semi-quantitative assessment, also attributed to the diversity of causes of ground glass radiographic appearances [15].

ROC analysis was employed to evaluate the performance of the automated 3D quantification scheme in detecting ILD patterns. While reliability analysis reflects the agreement of the 3D automated ILD quantification scheme with radiologist semi-quantitative assessment, ROC analysis is an additional performance evaluation metric reflecting the accuracy of the CAD scheme in detecting ILD patterns for all possible operating thresholds (cutoff points). The CAD scheme demonstrated a high ability in differentiating abnormal lung parenchyma from normal lung tissue and further identifying subtle ILD signs. The achieved performance ($A_z=0.950$) is comparable or superior to previously reported texture-based pixel/voxel-

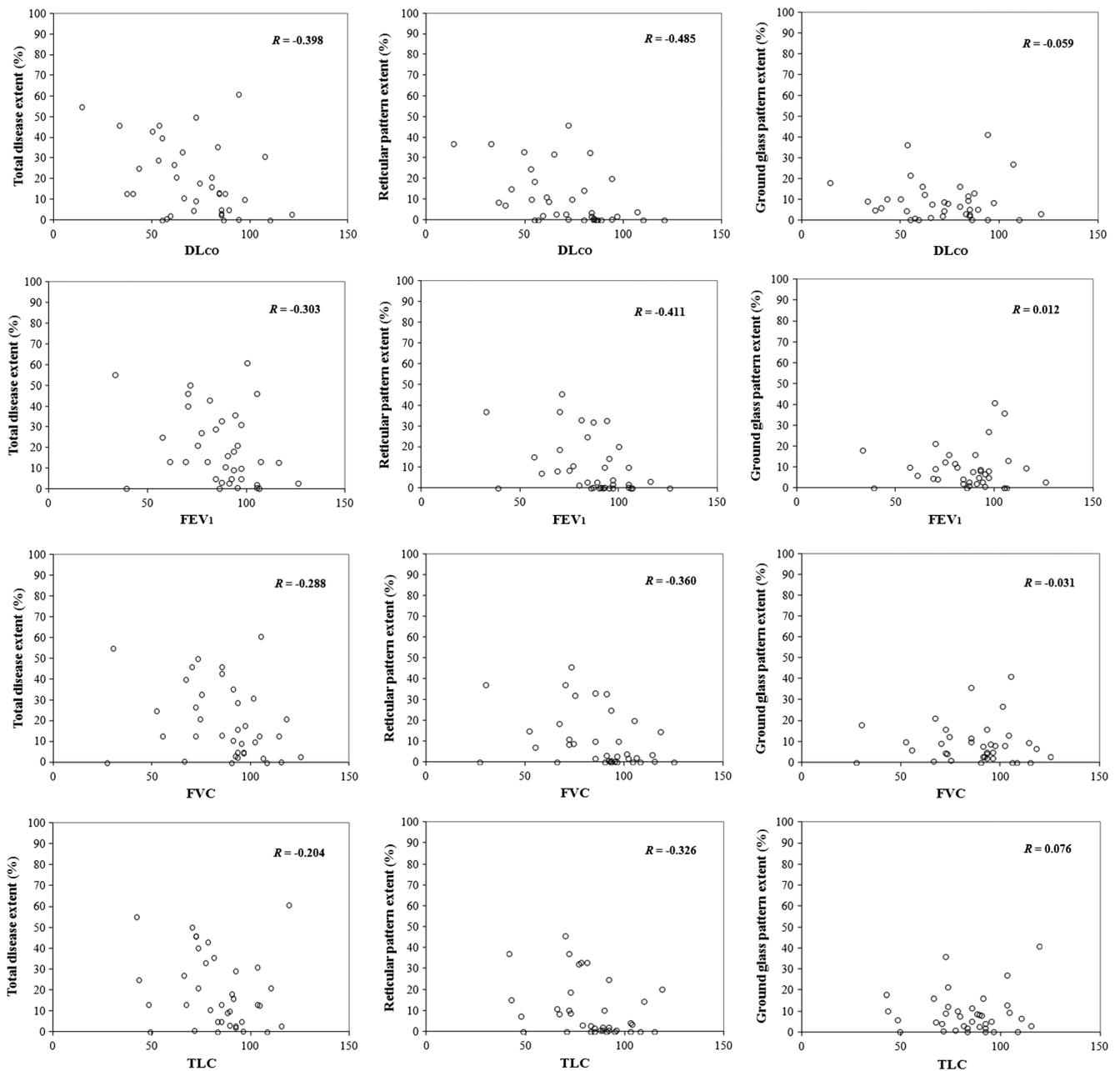


Fig. 4 Correlation between ILA extent semi-quantitatively assessed by the two radiologists in consensus (RAD_{cons}) and PFTs (FVC, FEV₁, TLC, and DL_{CO})

classification ILA quantification systems [18,19,21]. Specifically, Kim et al. [19] reported an A_z index equal to 0.86 in a subset of the originally analyzed HRCT dataset corresponding to agreed-on cases by radiologists. The ILA detection scheme presented in Park et al. [18] demonstrated accuracy of $A_z=0.884$ in a case sample of 30 HRCT patient scans. Similar performance ($A_z=0.885$) was achieved by an automated ILA quantification tool, reported by Rosas et al. [21] in a dataset of 86 HRCT patient scans.

The current study has also evaluated performance of the CAD scheme in detecting distinct reticular and ground glass

patterns, not, however, considered in previously reported ILA quantification schemes. Results have suggested that the CAD scheme has the ability to accurately detect distinct reticular and ground glass patterns. Furthermore, subtle signs of reticular and ground glass patterns may be captured and correctly identified by the CAD scheme. Specifically, the CAD scheme demonstrated a high performance in detecting reticular patterns ($A_z=0.920$), suggesting its potential in capturing tissue alterations reflecting disease progression. The high performance achieved by the CAD scheme in detecting ground glass patterns ($A_z=0.883$) may further contribute in improving

treatment decision making, since ground glass pattern partly corresponds to reversible inflammatory lung damage.

Toward further establishing validity of the presented ILD quantification scheme, the correlation between automatically derived disease extent and PFTs was investigated. Automatically derived total disease extent, as well as reticular pattern extent, correlates to specific PFT indexes, while ground glass pattern extent did not demonstrate correlation to any of the PFTs considered, suggesting that ground glass does not contribute in lung dysfunction. Similar trends with respect to correlation to PFTs are observed in case of disease extent assessed semi-quantitatively. These results also suggest a CAD scheme with high performance in quantifying disease extent.

The importance of investigating correlation of distinct ILD patterns with clinical parameters has been recently highlighted [9]. Such analysis is expected to provide insight into the relationship between specific CT finding and underlying pathophysiology and disease progression. However, to the best of the authors' knowledge, correlation of automatically derived distinct reticular and ground glass patterns to PFTs has not been investigated. Previously reported CAD-based ILD quantification studies have investigated correlation of total disease extent with PFTs [15–17] and reticular pattern extent to PFTs [19,21] demonstrating results in accordance to the ones presented herein, without, however, considering the contribution of distinct ground glass pattern. Specifically, Marten et al. [16] reported statistically significant negative correlation of automatically derived total disease extent to FVC (−0.586) and DL_{CO} (−0.583). Shin et al. [17] reported correlation of total disease extent quantified by a threshold-based scheme with DL_{CO} (−0.41). In Kim et al. [19], statistically significant negative correlation was found between CAD-based reticular disease extent and pulmonary function measurements of FVC (−0.31), TLC (−0.34), DL_{CO} (−0.35), and FEV_1 (−0.23). In Rosas et al. [21], reticular pattern extent assessed by a texture-based automated quantification demonstrated correlation to FVC and DL_{CO} (−0.483, −0.532 respectively).

Results of the current study are encouraging suggesting a reliable and accurate automated ILD quantification scheme. Optimizing detection accuracy in case of ground glass pattern and investigating performance evaluation in case of additional ILD patterns (e.g., consolidation) may further enhance potential of the automated ILD quantification scheme. Investigating correlation between CAD-quantified disease extent of varying distinct ILD patterns with clinical factors may cast further insight toward identification of disease underlying pathology manifested with specific imaging findings. Finally, investigating the feasibility of the computerized ILD quantification scheme in monitoring disease progression on follow-up image data accounts for an ongoing research effort.

Conclusions

The computerized ILD quantification scheme considered in the current study provides accurate detection and volumetric quantification of distinct reticular and ground glass patterns. Specifically, the automatically derived 3D quantification of disease extent is in agreement to radiologists' semi-quantitative scoring and correlates to PFTs. Computer-derived disease extent may potentially be used as an objective biomarker for ILD staging and treatment decision making in the clinical practice.

Acknowledgments This work is supported in part by the Caratheodory Programme (C.591) of the University of Patras, Greece.

References

1. Wells AU: High-resolution computed tomography and scleroderma lung disease. *Rheumatology* 47(Suppl 5):v59–v61, 2008
2. Wells AU, Behr J, Silver R: Outcome measures in the lung. *Rheumatology* 47(Suppl 5):v48–v50, 2008
3. Madani G, Katz RD, Haddock JA, Denton CP, Bell JR: The role of radiology in the management of systemic sclerosis. *Clin Radiol* 63(9):959–967, 2008
4. Goh NS, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al: Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med* 177(11):1248–1254, 2008
5. Wells AU, Steen V, Valentini G: Pulmonary complications: one of the most challenging complications of systemic sclerosis. *Rheumatology* 48(Suppl 3):iii40–iii44, 2009
6. Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al: Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med* 354(25):2655–2666, 2006
7. Desai SR, Veeraraghavan S, Hansell DM, Nikolakopoulou A, Goh NS, Nicholson AG, et al: CT features of lung disease in patients with systemic sclerosis: comparison with idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia. *Radiology* 232(2):560–567, 2004
8. Kazerooni EA, Martinez FJ, Flint A, Jamadar DA, Gross BH, Spizamy DL, et al: Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *AJR Am J Roentgenol* 169(4):977–983, 1997
9. Assayag D, Kaduri S, Hudson M, Hirsch A, Baron M: High resolution computed tomography scoring systems for evaluating interstitial lung disease in systemic sclerosis patients. *Rheumatology* S1:003, 2012. doi:10.4172/2161-1149.S1-003, 2012
10. Collins CD, Wells AU, Hansell DM, Morgan RA, MacSweeney JE, du Bois RM, et al: Observer variation in pattern type and extent of disease in fibrosing alveolitis on thin section computed tomography and chest radiography. *Clin Radiol* 49(4):236–240, 1994
11. Aziz ZA, Wells AU, Hansell DM, Bain GA, Copley SJ, Desai SR, et al: HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 59(6):506–511, 2004
12. Camiciottoli G, Orlandi I, Bartolucci M, Meoni E, Nacci F, Diciotti S, et al: Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. *Chest* 131(3):672–681, 2007
13. Lynch DA: Quantitative CT of fibrotic interstitial lung disease. *Chest* 131:643–654, 2007
14. Mascalchi M, Diciotti S, Sverzellati N, Camiciottoli G, Ciccotosto C, Falaschi F, Zompatori M: Low agreement of visual rating for detailed

- quantification of pulmonary emphysema in whole-lung CT. *Acta Radiol* 53(1):53–60, 2012
15. Marten K, Dicken V, Kneitz C, Höhmann M, Kenn W, Hahn D, et al: Interstitial lung disease associated with collagen vascular disorders: disease quantification using a computer-aided diagnosis tool. *Eur Radiol* 19(2):324–332, 2008
 16. Marten K, Dicken V, Kneitz C, Höhmann M, Kenn W, Hahn D, et al: Computer-assisted quantification of interstitial lung disease associated with rheumatoid arthritis: preliminary technical validation. *Eur J Radiol* 72(2):278–283, 2009
 17. Shin KE, Chung MJ, Jung MP, Choe BK, Lee KS: Quantitative computed tomographic indexes in diffuse interstitial lung disease: correlation with physiologic tests and computed tomography visual scores. *J Comput Assist Tomogr* 35(2):266–271, 2011
 18. Park SC, Tan J, Wang X, Lederman D, Leader JK, Kim SH, Zheng B: Computer-aided detection of early interstitial lung diseases using low-dose CT images. *Phys Med Biol* 56(4):1139–1153, 2011
 19. Kim HG, Tashkin DP, Clements PJ, Li G, Brown MS, Elashoff R, et al: A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol* 28(5, Suppl 62):S26–S35, 2010
 20. Korfiatis PD, Karahaliou AN, Kazantzi AD, Kalogeropoulou C, Costaridou L: Texture-based identification and characterization of interstitial pneumonia patterns in lung multidetector CT. *IEEE Trans Inf Technol Biomed* 14(3):675–680, 2010
 21. Rosas IO, Yao J, Avila NA, Chow CK, Gahl WA, Gochuico BR: Automated quantification of HRCT findings in individuals at risk for pulmonary fibrosis. *Chest* 140(6):1590–1597, 2011
 22. Depeursinge A, Vargas A, Gaillard F, Platon A, Geissbuhler A, Poletti PA, Müller H: Case-based lung image categorization and retrieval for interstitial lung diseases: clinical workflows. *Int J Comput Assist Radiol Surg* 7(1):97–110, 2012
 23. Ko JP, Naidich DP: Computer-aided diagnosis and the evaluation of lung disease. *J Thorac Imaging* 19(3):136–155, 2004
 24. Johkoh T, Müller NL, Nakamura H: Multidetector spiral high-resolution computed tomography of the lungs: distribution of findings on coronal image reconstructions. *J Thorac Imaging* 17(4):291–305, 2002
 25. Beigelman-Aubry C, Hill C, Guibal A, Savatovsky J, Grenier PA: Multi-detector row CT and postprocessing techniques in the assessment of diffuse lung disease. *Radiographics* 25(6):1639–1652, 2005
 26. Uchiyama Y, Katsuragawa S, Abe H, Shiraishi J, Li F, Li Q, Zhang CT, Suzuki K, Doi K: Quantitative computerized analysis of diffuse lung disease in high-resolution computed tomography. *Med Phys* 30(9):2440–2454, 2003
 27. Sluimer IC, Prokop M, Hartmann I, van Ginneken B: Automated classification of hyperlucency, fibrosis, ground glass, solid, and focal lesions in high-resolution CT of the lung. *Med Phys* 33(7):2610–2620, 2006
 28. Uppaluri R, Hoffman EA, Sonka M, Hartley PG, Hunninghake GW, McLennan G: Computer recognition of regional lung disease patterns. *Am J Respir Crit Care Med* 160(2):648–654, 1999
 29. Zavaletta VA, Bartholmai BJ, Robb RA: High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Acad Radiol* 14(7):772–787, 2007
 30. Xu Y, van Beek EJ, Hwanjo Y, Guo J, McLennan G, Hoffman EA: Computer-aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM). *Acad Radiol* 13(8):969–978, 2006
 31. Sørensen L, Shaker SB, de Bruijne M: Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Trans Med Imaging* 29(2):559–569, 2010
 32. American College of Rheumatology. <http://www.rheumatology.org/>. Accessed 1 March 2013
 33. Bongartz G, Golding SJ, Jurik AG, et al: European guidelines for quality criteria for computed tomography. EUR 16262. The European Commission's Study Group on Development of Quality Criteria for Computed Tomography. European Commission, Luxembourg, 2000
 34. Korfiatis P, Kalogeropoulou C, Karahaliou A, Kazantzi A, Skiadopoulos S, Costaridou L: Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. *Med Phys* 35(12):5290–5302, 2008
 35. Korfiatis P, Skiadopoulos S, Sakellaropoulos P, Kalogeropoulou C, Costaridou L: Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin-slice CT. *Br J Radiol* 80(960):996–1004, 2007
 36. Korfiatis PD, Kalogeropoulou C, Karahaliou AN, Kazantzi AD, Costaridou LI: Vessel tree segmentation in presence of interstitial lung disease in MDCT. *IEEE Trans Inf Technol Biomed* 15(2):214–220, 2011
 37. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al: ATS/ERS Task Force. Standardisation of spirometry. *Eur Respir J* 26(2):319–338, 2005
 38. Anonymous: Standardized lung function testing. Official statement of the European Respiratory Society. *Eur Respir J Suppl* 16:1–100, 1993
 39. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychological Bull* 86(2):420–428, 1979
 40. Erkel AR, Pattynama PM: Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *Eur J Radiol* 27(2):88–94, 1998