CHEST

Quantitative computed tomography assessment for systemic sclerosis–related interstitial lung disease: comparison of different methods

Furkan Ufuk[1] · Mahmut Demirci[1] · Goksel Altinisik[2]

Abstract

Objectives To compare the previously defined six different histogram-based quantitative lung assessment (QLA) methods on high-resolution CT (HRCT) in patients with systemic sclerosis (SSc)–related interstitial lung disease (ILD).

Methods The HRCT images of SSc patients with ILD were reviewed, and the visual ILD score (semiquantitative) and the severity of ILD (limited or extensive) were calculated. The QLA score of ILD was evaluated using the previously defined six different methods and parameters (different lung attenuation ranges, skewness, kurtosis, mean lung attenuation, and standard deviation [SD]). Pulmonary function tests (PFTs) were also performed on all patients. Relationships among variables were evaluated using Spearman's correlation coefficient (r). Diagnostic performance of quantitative methods for the ability to differentiate the limited from extensive ILD was calculated using ROC analysis.

Results Fifty-five patients were included in the study. There was a significant correlation between all quantitative and semiquantitative measurement results ($p < 0.0001$). The QLA scores revealed a significant correlation with PFT results. The kurtosis value of the voxels between -200 and -1024 Hounsfield unit (HU) (Method-5) showed the best correlation with semiquantitative evaluation ($r = -0.740$, $p < 0.0001$). The ROC analysis demonstrated the best performance of SD of the voxels between -400 and -950 HU (Method-6) for histogram analysis method and Method-3 (voxels between -260 and -600 HU were calculated as ILD) for CT density cutoff methods.

Conclusions All the QLA methods are applicable in assessing the ILD score in SSc patients and have potential importance to differentiate limited from extensive ILD.

Key Points

• *Quantitative interstitial lung disease assessment helps clinicians to assess systemic sclerosis patients with interstitial lung disease.*

• *Quantitative lung assessment methods are applicable in assessing the interstitial lung disease score in systemic sclerosis patients.*

• *Quantitative lung assessment methods have potential importance in the management of patients.*

Keywords Multidetector computed tomography · Interstitial lung diseases · Pulmonary function tests · Quantitative evaluation · Systemic sclerosis

---

✉ Furkan Ufuk
  furkan.ufuk@hotmail.com

1   Department of Radiology, School of Medicine, University of Pamukkale, Kinikli, 20100 Denizli, Turkey

2   Department of Chest Diseases, School of Medicine, University of Pamukkale, Kinikli, 20100 Denizli, Turkey

Abbreviations

AUC	Area under the curve
CII	Computerized integrated index
CII-5	Computerized integrated index of Method-5
CII-6	Computerized integrated index of Method-6
CT	Computed tomography
CVDs	Collagen vascular diseases
DLCO	Single-breath diffusing capacity
DcSSc	Diffuse forms of systemic sclerosis
FEV1	Forced expiratory volume in 1 s

FVC	Forced vital capacity
HRCT	High-resolution computed tomography
HU	Hounsfield unit
ILD	Interstitial lung disease
KURT-5	Kurtosis value of Method-5
KURT-6	Kurtosis value of Method-6
LcSSc	Limited forms of systemic sclerosis
MLA	Mean lung attenuation
MLA-5	Mean lung attenuation value of Method-5
MLA-6	Mean lung attenuation value of Method-6
PCA	Principal component analysis
PFT	Pulmonary function test
PRoTA	Pattern recognition or texture analysis
QLA	Quantitative lung assessment
QUANT-1	Quantitative Method-1
QUANT-2	Quantitative Method-2
QUANT-3	Quantitative Method-3
QUANT-4	Quantitative Method-4
ROC	Receiver operating characteristic
SD	Standard deviation
SD-5	Standard deviation value of Method-5
SD-6	Standard deviation value of Method-6
SKEW-5	Skewness value of Method-5
SKEW-6	Skewness value of Method-6
SSc	Systemic sclerosis
TLC	Total lung capacity

Introduction

Systemic sclerosis (SSc, also known as systemic scleroderma) is a systemic disorder with autoimmune features and characterized by fibrosis of the visceral organs and skin, which may also affect musculoskeletal, gastrointestinal, respiratory, and cardiovascular systems [1]. Interstitial lung disease (ILD) is frequently seen in SSc patients, and ILD accounts for 33% of deaths in these patients [2, 3]. High-resolution computed tomography (HRCT) is one of the most widely used imaging methods and mandatory for the evaluation of ILD in SSc patients. It has also been reported that HRCT findings in ILD (particularly severity of ILD) significantly correlate with pulmonary function tests (PFTs), which have a prognostic value for ILDs [4, 5]. However, accurate interpretation of HRCT images and the semiquantitative assessment of ILD are still a problem for inexperienced general radiologists, rheumatologists, and pulmonologists as there is still wide intra- and inter-observer variability even among experienced thoracic radiologists [6, 7]. In addition, the wide inter-observer variability may be a consequence of the ILD diversity. Furthermore, the diagnosis of progression or regression of the disease in patients with ILD has prognostic significance and determines whether to continue or change treatment [2, 5]. Therefore, a quantitative (objective), noninvasive, reproducible, and reliable method that allows for the accurate evaluation of ILD is highly desirable.

To date, several quantitative lung assessment (QLA) methods to investigate lung involvement in ILDs have been developed. These include pattern recognition or texture analysis (PRoTA) methods and histogram-based methods [8–26]. Pattern recognition or texture analysis methods are more complex and usually need an additional application or software. Moreover, it has been demonstrated that PRoTA has a weaker correlation with PFT results and weaker visual assessment scores than the histogram-based methods [19–26].

Histogram-based QLA methods, which are based on the characterization of lung tissue by using a histogram analysis, are simpler, freely accessible, and commonly used [8–18]. The studies using histogram analysis methods can be classified under two main methods. The first is the computed tomography (CT) histogram analysis method, which includes multiple parameters (such as skewness, kurtosis, mean lung attenuation values, and standard deviations [SDs]) [15–18]. The second is the CT density (Hounsfield unit [HU]) cutoff or HU percentile method, which is based on the detection of pathological ILD areas or volumes in the CT histogram analyses [8–14]. To date, six different histogram-based QLA methods and threshold values have been used for the detection of pathological ground-glass or fibrosis areas [8–18]. To our knowledge, there is no standard CT threshold value defined for the assessment of ILD severity in the CT histogram analyses. Therefore, our aim is to compare the previously defined six different histogram-based QLA methods and threshold values in SSc patients with ILDs.

Materials and methods

Our study was approved by the Ethics Committee on Human Experiments of Pamukkale University, with approval number 60116787-020/39922.

Study population

We retrospectively investigated consecutive patients with known SSc and associated ILD who underwent thin-section HRCT and PFTs from February 2016 to March 2019 for the first evaluation or follow-up. The SSc diagnosis was made using the 2013 classification criteria for SSc [27]. The inclusion criteria for HRCT analysis were that patients with SSc-associated ILD had PFTs within 2 weeks after or before undergoing HRCT. Exclusion criteria were the presence of pulmonary edema, infection, or lung mass, known moderate to severe pulmonary arterial hypertension with a mean pulmonary artery pressure ≥ 30 mmHg, and the presence of major motion artifacts, which may affect assessment of HRCT. In addition, patients who were unable to cooperate with the PFTs

were excluded. Patients were classified into limited (lcSSc) and diffuse (dcSSc) forms of SSc. Skin involvement in lcSSc form is limited to the face, hands, feet, forearms, and lower legs. However, skin thickening in the dcSSc form may additionally include upper arms, thighs, and trunk. [27].

Pulmonary function tests

Pulmonary function tests were performed within 2 weeks from the HRCT examinations. Forced vital capacity (FVC), forced expiratory volume in 1 s (FEV1), total lung capacity (TLC), and single-breath diffusing capacity (DLCO) were measured. The PFTs were obtained by using commonly accepted techniques, and results were expressed as a percentage of predicted performance [28]. A minimum of three measurements were taken for each variable to guarantee repeatability.

HRCT examinations

High-resolution computed tomography images were obtained from the lung apices to the bases using a multidetector CT system (Ingenuity Core 128, Philips Medical Systems) at full inspiration in the supine position. The parameters were 64×0.625 mm slice collimation, 1.5-mm-slice thickness and 0.75 mm reconstructions, 0.4-s rotation time, 250–300 mm field of view, 100 kV tube voltage, and 250–300 mA tube current. The HRCT images were reconstructed using a high-spatial-frequency algorithm. Intravenous contrast medium was not used. A 32-cm-diameter phantom was used to represent an adult's body and the mean $CTDI_{vol}$ and DLP values were 7.2 mGy and 245.9 mGy cm, respectively.

Semiquantitative image analysis

All HRCT images were reviewed at lung window settings (with a window center of -500 to -600 HU and a window width of 1600 HU) by two observers in consensus who were unaware of the clinical findings or PFT results.

First, HRCT images were evaluated for suitability for the study, and non-eligible patients, according to the exclusion criteria, were excluded. Then, all HRCT images were evaluated according to the ILD staging system, which was described by Goh et al [29] (Supplementary; Fig. 1). Briefly, CT images were scored at five levels (the origin of great vessels, the main carina, the pulmonary venous confluence, halfway between the third and fifth sections, and immediately above the right hemidiaphragm) and the disease extent was estimated as a percentage of the total area to the nearest 5% in each of the five CT sections. The total extent of ILD (TEI) detected by HRCT was calculated as the mean extent score in the five scored CT sections. The TEI $\leq 10\%$ was accepted as limited, and $\geq 30\%$ was accepted as an extensive disease. For indeterminate cases (with the extent of ILD at 10–30%), FVC

was considered as FVC $\geq 70\%$ indicating limited disease and FVC $< 70\%$ indicating extensive disease [29]. We also calculated the modified coarseness of reticular disease (MCRD) in each of the five sections according to Goh et al [29] as follows: 0, normal; 1, ground-glass opacity alone; 2, fine intra-lobular fibrosis; 3, microcystic honeycombing (≤ 4 mm); 4, macrocystic honeycombing (> 4 mm). The highest score in each section was taken into account and the MCRD was the summed score for all five levels and MCRD accepted as ≤ 10 (mild reticulation) and > 10 (severe reticulation).

Quantitative image analysis

All HRCT images were analyzed by a single trained radiologist using a free DICOM viewer software (Osirix Version 8.0; Pixmeo SARL). For each slice, a semiautomatic segmentation of the lung parenchyma was performed in order to obtain an analysis of whole CT images. Afterward, the descriptive parameters of the quantitative analysis were calculated in different CT attenuation values, as previously described in detail [8–13, 15, 16, 18]. All histogram-based quantitative assessments were made in six different sessions by the trained radiologist (Table 1; Fig. 2). Due to the nature of quantitative measurement, agreement between observers has not been investigated. If necessary, minimal user intervention was performed to exclude pulmonary vessels, esophagus, trachea, and main bronchi.

Statistical analysis

Continuous data represented as a mean with an SD or a median with a range and categorical variables were represented as a percentage (%). The Shapiro-Wilk W test was used for assessing the normality of the data, and appropriate tests were selected accordingly. Relationships among variables were evaluated using Spearman's correlation coefficient (r). A Spearman r value of 0–0.30 was considered fair, 0.31–0.50 moderate, 0.51–0.70 good, and 0.71–1.00 excellent correlation. Student's t test and the chi-square test were used to compare continuous and categorical variables, respectively. The mean lung attenuation (MLA), SD, skewness, and kurtosis values of Method-5 and Method-6 were combined, and computerized integrated index (CII) values were calculated using principal component analysis (PCA), which is a complex mathematical technique that transforms a number of data sets (in this case, MLA, skewness, and kurtosis) into single variable (CII) with a minimal loss of information [30]. Using this method, CII-5 value and CII-6 value were calculated separately. The Bland-Altman analysis was used to investigate the agreement between quantitative methods. To calculate the diagnostic performance of QLA methods, including CII-5 and CII-6 values, for the diagnosis of severity of ILD (limited or extensive), receiver operating characteristic (ROC) analysis
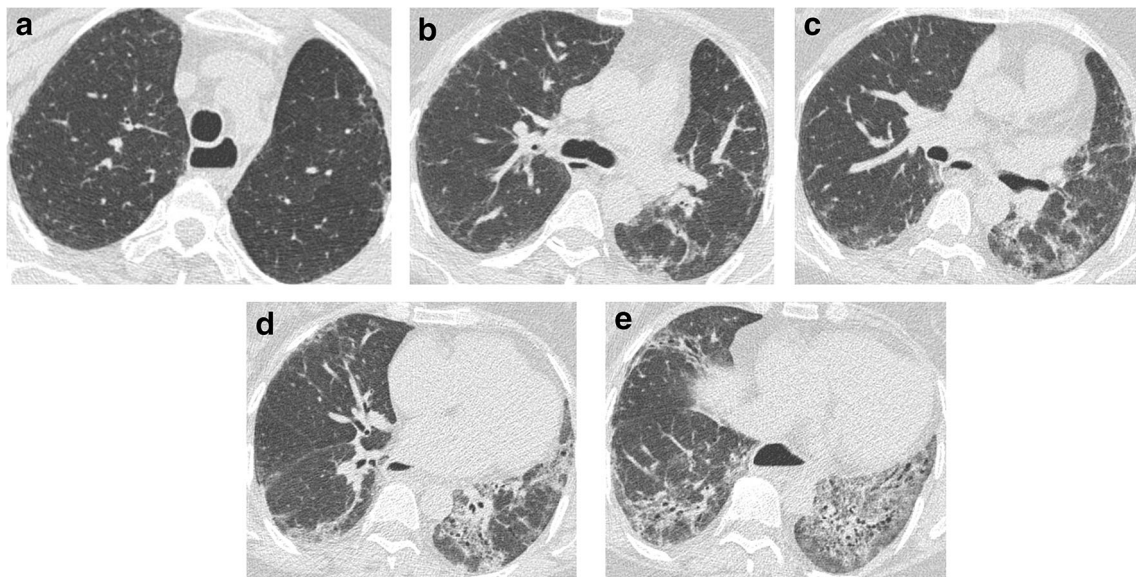
Fig. 1 An example of a semiquantitative high-resolution computed tomography evaluation. High-resolution computed tomography images at five levels. **a** The origin of great vessels (estimated disease extent score was 10% and modified coarseness of reticular disease was 2). **b** The main carina (15% and 2). **c** The pulmonary venous confluence (40% and 1). **d** Halfway between the third and fifth sections (55% and 1). **e** Immediately above the right hemidiaphragm (65% and 1). The estimated mean interstitial lung disease extent score was 37% and the estimated mean coarseness of reticular disease was 7

was used. A *p* value of less than 0.05 was considered statistically significant. The statistical analysis was performed on a personal computer using SPSS (Version 24.0, IBM) and MedCalc (Version 16.1, MedCalc Software).

Results

A total of 55 patients (45 female) with SSc-related ILD were included in the study (Fig. 3). Males were significantly taller

and weightier than females ($p = 0.001$ and $p = 0.021$, respectively). No significant differences of FVC, FEV1, DLCO, TLC, and QLA scores were found between sexes. The characteristics of the patients (24 with lcSSc, 31 with dcSSc) and the PFT results are shown in Table 2.

Semiquantitative assessment results

The average disease extent on the semiquantitative assessment was $17.9 \pm 14.7\%$ (range 1–62%). When we investigated the

**Table 1** Histogram-based quantitative lung assessment methods

Method	Description
Method-1 (QUANT-1)	To calculate lung volume, all voxels between − 200 and − 1024 HU were selected. All voxels between − 500 and − 700 HU were calculated as ILD, which was described by Salaffi et al [10, 11]. The quantitative assessment score (Quant-1) was calculated as (quantitative ILD volume / quantitative total lung volume) × 100.
Method-2 (QUANT-2)	To calculate lung volume, all voxels between − 250 and − 1024 HU were selected. All voxels between − 500 and − 800 HU were calculated as ILD, which was described by Yabuuchi et al [9]. The quantitative assessment score (Quant-2) was calculated as (quantitative ILD volume / quantitative total lung volume) × 100.
Method-3 (QUANT-3)	To calculate lung volume, all voxels between − 200 and − 1024 HU were selected. All voxels between − 260 and − 600 HU were calculated as ILD, which was described by Ninaber et al [8]. The quantitative assessment score (Quant-3) was calculated as (quantitative ILD volume / quantitative total lung volume) × 100.
Method-4 (QUANT-4)	To calculate lung volume, all voxels between − 400 and − 1024 HU were selected. All voxels between − 400 and − 800 HU were calculated as ILD, which was described by Marten et al [12, 13]. The quantitative assessment score (Quant-4) was calculated as (quantitative ILD volume / quantitative total lung volume) × 100.
Method-5	To isolate the lungs, all voxels between − 200 and − 1024 HU were selected. Values for mean lung attenuation (MLA), SD, skewness, and kurtosis of the whole lung parenchyma were calculated, which was described by Koyama et al [18].
Method-6	To isolate the lungs, all voxels between − 400 and − 950 HU were selected. Values for MLA, SD, skewness, and kurtosis of the whole lung parenchyma were calculated, which was described by Ariani et al [15, 16].
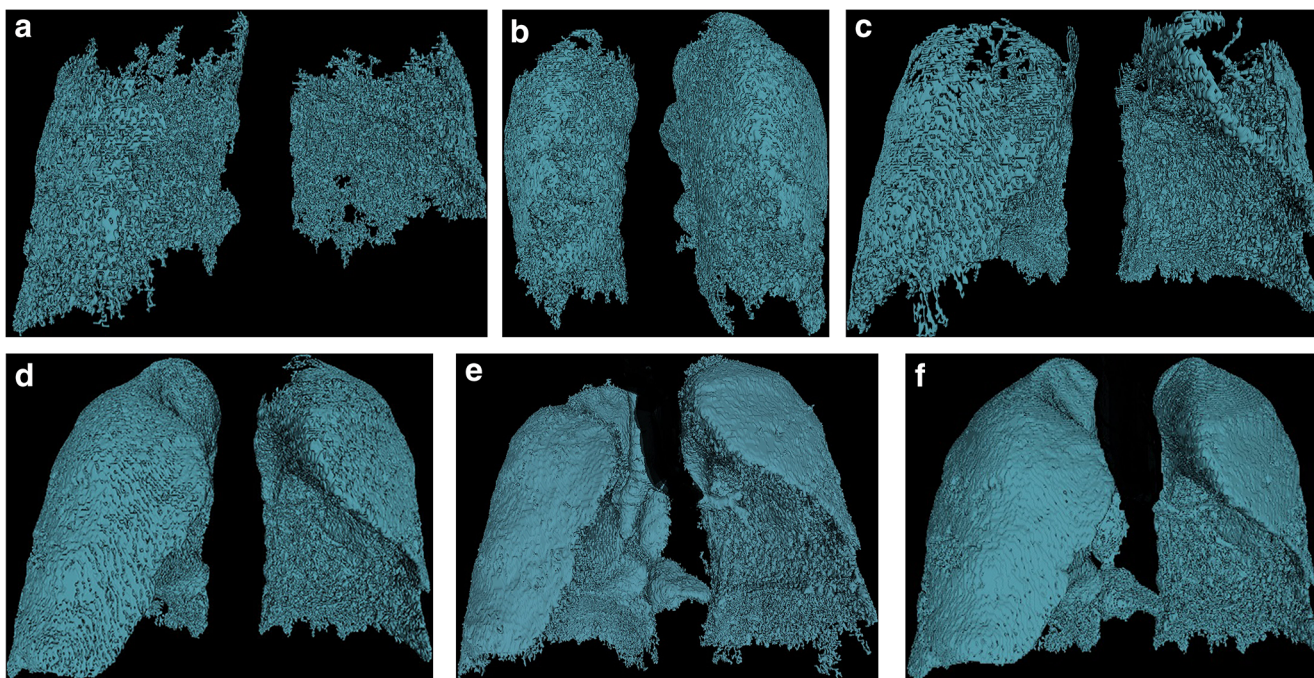
Fig. 2 An example of a quantitative high-resolution computed tomography evaluation. **a** All voxels between −500 and −700 HU for Method-1 (volume 191.6 cm^3). **b** All voxels between −500 and −800 HU for Method-2 (volume 646.7 cm^3). **c** All voxels between −260 and −600 HU for Method-3 (volume 314.9 cm^3). **d** All voxels between −400 and −800 HU for Method-4 (volume 779.1 cm^3). **e** All voxels between −200 and −1024 HU for Method-5 (volume 1852.7 cm^3, mean lung attenuation −719.4 HU, standard deviation 204.7, skewness 0.863, kurtosis −0.144). **f** All voxels between −400 and −950 HU for Method-6 (volume 1497.7 cm^3, mean lung attenuation −753.2 HU, standard deviation 135.8, skewness 0.795, kurtosis −0.215)

Fig. 3 Patient selection and inclusion flow diagram of systemic sclerosis patients with interstitial lung disease
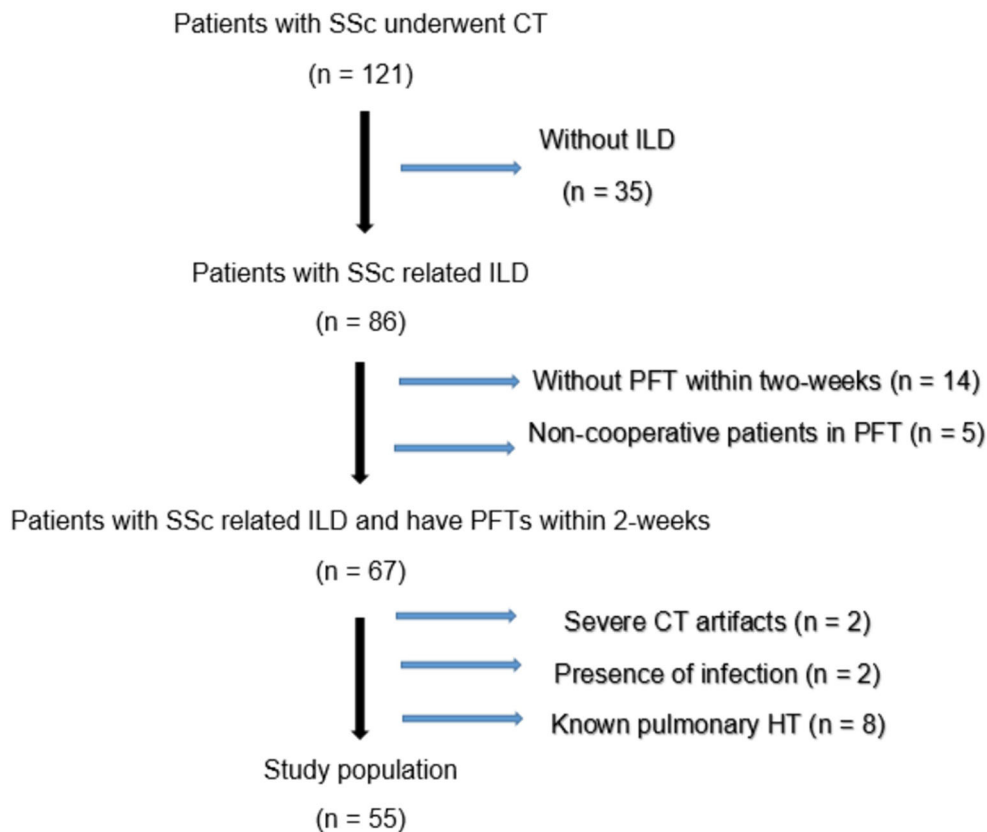
disease extent using a method that was described by Goh et al [29], 37 patients (67.3%) had limited and 18 patients (22.7%) had extensive ILD. The mean total coarseness score was 8.4 ± 3.5 (range 3–16). When we investigated the coarseness of reticular disease, 44 (80%) patients had mild and 11 (20%) had severe reticulation.

Correlation between semiquantitative results and PFTs

The semiquantitative disease extent score, which was assessed by two observers in consensus, revealed a good negative correlation with FVC ($r = -0.593$, $p < 0.0001$) and FEV1 ($r = -0.553$, $p < 0.0001$). The semiquantitative disease extent score showed a moderate negative correlation with TLC ($r = -0.494$, $p < 0.0001$) and the DLCO results ($r = -0.398$, $p = 0.005$). A statistically significant relationship was found between coarseness of reticular disease and FVC ($p = 0.015$, $r = -0.326$). There was no significant relationship between coarseness of reticular disease and DLCO ($p = 0.67$) or FEV1 ($p = 0.184$) values.

Correlation and comparison of quantitative methods by CT density cutoff methods

The average disease extent on the quantitative assessment according to Method-1, -2, -3, and -4 was 27.3 ± 15.3% (range 8.5–67.2%), 36 ± 18.5% (range 5.2–79.4%), 11.4 ± 10.1% (range 0.2–39.1%), and 43 ± 21.3% (range 9.1–85.4%), respectively. The mean disease extent score calculated by different QLA methods had a wide range of distribution. These can be due to overestimation (due to the identification of pulmonary vasculature as ILD; Method-1, -2, -4) or underestimation (due to incomplete lung segmentation; Method-3) of ILD in QLA methods, as described before [10–13]. Therefore, the

user should be aware of the possibility of overestimation and underestimation in these QLA methods. Although the difference in extent assessment between methods is quite wide, there was a close correlation between QLA methods ($p < 0.0001$) (Table 3).

Correlation between semiquantitative and quantitative results

There was a close correlation between QLA methods and semiquantitative (visual) measurement results ($p < 0.0001$). The kurtosis value of Method-5 (KURT-5; kurtosis value of the voxels between − 200 and − 1024 HU) showed the maximum correlation with semiquantitative evaluation ($r = -0.740$, $p < 0.0001$) (Table 4). The average difference between the semiquantitative assessment and quantitative methods is also an estimate of agreement, and it was found to be very small (Fig. 4).

An ROC analysis was performed for the QLA methods for discrimination of limited and extensive ILD (Fig. 5). The ROC analysis demonstrated the best and excellent performances of the SD of Method-6 (SD-6; SD value of the voxels between − 400 and − 950 HU) for histogram analysis methods (sensitivity and specificity for the cutoff value of 119.5 HU were 88.9% and 80.6%, respectively), and Method-3 for CT density cutoff methods (sensitivity and specificity for cutoff values of 11.3% extensity score were 83.3% and 75%, respectively), with an area under the curve (AUC) of 0.903 and 0.873, respectively (Table 5). The CII-5 and CII-6 explained 84.5% and 89% of the total variability, respectively. CII values showed a negative correlation with MLA ($r = -0.91$ and − 0.92 for CII-5 and CII-6, respectively), and positive correlation with skewness ($r = 0.95$ and 0.98, respectively) and kurtosis ($r = 0.97$ and 0.98, respectively). The discrimination of CII-5 (the combination of MLA, skewness, and kurtosis

Table 2 The characteristics of the patients and pulmonary function tests results

	Total population Mean ± SD (range)	Females Mean ± SD	Males Mean ± SD	*p* value
Age (years)	55.6 ± 13.9 (30–81)	53.9 ± 13.7	63.2 ± 12.9	0.06
Height (cm)	158 ± 7.1 (145–177)	156.3 ± 6.2	165.5 ± 6.2	0.001
Weight (kilograms)	64.4 ± 11.9 (44–94)	62.1 ± 10.5	74.3 ± 13.5	0.021
Body mass index (kg/m^2)	25.8 ± 4.5	25.5 ± 4.5	27.1 ± 4.3	0.319
Duration of SSc (years)	2.6 ± 1.5 (0.03–7.8)	2.4 ± 1.5	3.6 ± 1.3	0.624
FVC (% of predicted)	84.8 ± 26.7 (40–136.9)	87 ± 25.3	74.5 ± 31.7	0.265
FEV1 (% of predicted)	83.3 ± 24.5 (43–147.2)	85 ± 23.9	76 ± 27	0.350
Total lung capacity (% of predicted)	76.5 ± 25.8 (25–163)	77.8 ± 25.1	70.1 ± 29.5	0.530
DLCO (% of predicted)	67.4 ± 22.2 (25–163)	67.7 ± 19.9	66.1 ± 32.1	0.892
Mean time interval between HRCT and PFTs (days)	6.6 ± 3.3 (0–12)	6.5 ± 3.5	6.6 ± 4.2	0.744

SSc systemic sclerosis, *FVC* forced vital capacity, *FEV1* forced expiratory volume in 1 s, *DLCO* single-breath diffusing capacity, *HRCT* high-resolution computed tomography, *PFTs* pulmonary function tests

Table 3 Correlations between quantitative methods

		QUANT-1	QUANT-2	QUANT-3	QUANT-4	MLA-5	SD-5	SKEW-5	KURT-5	MLA-6	SD-6	SKEW-6	KURT-6	CII-5	CII-6
QUANT-1	r value	1	0.926**	0.967**	0.942**	0.977**	0.698**	−0.898**	−0.945**	0.975**	0.750**	−0.936**	−0.952**	−0.952**	−0.952**
	p value		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
QUANT-2	r value	0.926**	1	0.861**	0.967**	0.931**	0.464**	−0.800**	−0.802**	0.941**	0.511**	−0.855**	−0.835**	−0.835**	−0.835**
	p value	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
QUANT-3	r value	0.967**	0.861**	1	0.884**	0.944**	0.717**	−0.844**	−0.923**	0.922**	0.760**	−0.865**	−0.910**	−0.910**	−0.910**
	p value	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
QUANT-4	r value	0.942**	0.967**	0.884**	1	0.940**	0.532**	−0.821**	−0.838**	0.947**	0.585**	−0.879**	−0.864**	−0.864**	−0.864**
	p value	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
MLA-5	r value	0.977**	0.931**	0.944**	0.940**	1	0.657**	−0.830**	−0.894**	0.964**	0.658**	−0.878**	−0.897**	−0.897**	−0.897**
	p value	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
SD-5	r value	0.698**	0.464**	0.717**	0.532**	0.657**	1	−0.687**	−0.798**	0.628**	0.855**	−0.699**	−0.755**	−0.755**	−0.755**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
SKEW-5	r value	−0.898**	−0.800**	−0.844**	−0.821**	−0.830**	−0.687**	1	0.951**	−0.894**	−0.824**	0.963**	0.957**	0.957**	0.957**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
KURT-5	r value	−0.945**	−0.802**	−0.923**	−0.838**	−0.894**	−0.798**	0.951**	1	−0.912**	−0.867**	0.941**	0.964**	0.964**	0.964**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
MLA-6	r value	0.975**	0.941**	0.922**	0.947**	0.964**	0.628**	−0.894**	−0.912**	1	0.683**	−0.937**	−0.937**	−0.937**	−0.937**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001	0.0001
SD-6	r value	0.750**	0.511**	0.760**	0.585**	0.658**	0.855**	−0.824**	−0.867**	0.683**	1	−0.812**	−0.866**	−0.866**	−0.866**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001	0.0001
SKEW-6	r value	−0.936**	−0.855**	−0.865**	−0.879**	−0.878**	−0.699**	0.963**	0.941**	−0.937**	−0.812**	1	0.983**	0.983**	0.983**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001	0.0001
KURT-6	r value	−0.952**	−0.835**	−0.910**	−0.864**	−0.897**	−0.755**	0.957**	0.964**	−0.937**	−0.866**	0.983**	1	0.930**	0.930**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001	0.0001
CII-5	r value	−0.977	−0.879	−0.950	−0.903	−0.978	−0.782	0.857**	0.934**	−0.950	−0.754	0.906**	0.930**	1	0.960**
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001		0.0001
CII-6	r value	−0.982	−0.897	−0.940	−0.917	−0.946	−0.715	0.924**	0.952**	−0.981	−0.797	0.966**	0.977**	0.960**	1
	p value	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	

* indicates a *p* value < 0.01; ** indicates a *p* value < 0.05

QUANT-1 quantitative Method-1, *QUANT-2* quantitative Method-2, *QUANT-3* quantitative Method-3, *QUANT-4* quantitative Method-4, *MLA-5* mean lung attenuation value of Method-5, *SD-5* standard deviation value of Method-5, *SKEW-5* skewness value of Method-5, *KURT-5* kurtosis value of Method-5, *MLA-6* mean lung attenuation value of Method-6, *SD-6* standard deviation value of Method-6, *SKEW-6* skewness value of Method-6, *KURT-6* kurtosis value of Method-6, *CII-5* computerized integrated index of Method-5, *CII-6* computerized integrated index of Method-6

Table 4 Correlation between semiquantitative and quantitative results

		Quant-1	Quant-2	Quant-3	Quant-4	MLA-5	SD-5	SKEW-5	KURT-5	MLA-6	SD-6	SKEW-6	KURT-6	CII-5	CII-6
		Quantitative assessment methods													
FVC	r	−0.342*	−0.23	−0.418**	−0.209	−0.28	−0.382**	0.283*	0.327*	−0.243	−0.409**	0.314*	0.366**	0.302*	0.314**
	p	0.016	0.042	0.003	0.049	0.051	0.007	0.049	0.022	0.093	0.004	0.028	0.01	0.035	0.028
DLCO	r	−0.0384**	−0.303*	−0.397**	−0.311*	−0.320*	−0.392**	0.489**	0.507**	−0.378**	−0.490**	0.449**	0.470**	0.380**	0.449**
	p	0.004	0.024	0.003	0.021	0.017	0.003	0.0001	0.0001	0.004	0.0001	0.001	0.0001	0.004	0.001
FEV1	r	−0.357**	−0.291*	−0.0353**	−0.277*	−0.291*	−0.357**	0.459**	0.464**	−0.360**	−0.434**	0.433**	0.441**	0.345**	0.416**
	p	0.008	0.033	0.009	0.043	0.032	0.008	0.0001	0.0001	0.007	0.001	0.001	0.001	0.011	0.002
Visual score	r	0.626**	0.444**	0.619**	0.462**	0.587**	0.618**	−0.673**	−0.740**	0.569**	0.718**	−0.643**	−0.689**	−0.640	−0.645
	p	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

* indicates a p value < 0.01; ** indicates a p value < 0.05

QUANT-1 quantitative Method-1, QUANT-2 quantitative Method-2, QUANT-3 quantitative Method-3, QUANT-4 quantitative Method-4, MLA-5 mean lung attenuation value of Method-5, SD-5 standard deviation value of Method-5, SKEW-5 skewness value of Method-5, KURT-5 kurtosis value of Method-5, MLA-6 mean lung attenuation value of Method-6, SD-6 standard deviation value of Method-6, SKEW-6 skewness value of Method-6, KURT-6 kurtosis value of Method-6, CII-5 computerized integrated index of Method-5, CII-6 computerized integrated index of Method-6

values of Method-5) and CII-6 (the combination of MLA, skewness, and kurtosis values of Method-6) values for extensive ILD were 0.63 (AUC = 0.836; sensitivity 83.3%; specificity 72.2%) and 0.60 (AUC = 0.860; sensitivity 83.3%; specificity = 77.8%), respectively.

## Correlation between quantitative measurement results and PFTs

The QLA scores of all six methods demonstrated significant correlation with FVC, FEV1, and DLCO values. The best correlation with FVC value was found in Method-3 and the best correlation with DLCO and FEV1 values was found in the kurtosis value of Method-5 (Table 4).

Discussion

Herein, we investigated previously defined methods of QLA in SSc patients with ILD, and our results demonstrate a significant correlation between all QLA methods (both histogram analysis and CT density cutoff methods) ($p < 0.0001$). Moreover, all QLA methods reveal a significant correlation with semiquantitative (visual) assessment and PFTs (FVC, FEV1, and DLCO). All QLA methods were found to be successful in distinguishing between limited and extensive diseases that have a high prognostic impact. Although, SD-6 (standard deviation value of the voxels between −400 and −950 HU) was found to be the best discriminative parameter for histogram analysis methods, Method-3 demonstrated the best and excellent performance for CT density cutoff methods.

As mentioned, semiquantitative ILD staging system can provide prognostic information (limited or extensive disease) and follow-up in SSc patients with ILD [29]. Salaffi et al [10, 11] used a semiquantitative disease extent method which was described by Warrick and colleagues [31] and they found an excellent correlation between quantitative (Method-1) and semiquantitative disease extent scores ($r = 0.829$ and $r = 0.718$, respectively; $p < 0.0001$ for both). In our study, we found a good correlation between Method-1 and the semiquantitative disease extent score ($p < 0.0001$, $r = 0.626$). This small difference between the studies may be due to the use of different semiquantitative disease extent methods (Goh and colleagues [29] versus Warrick and colleagues [31]). Yabuuchi et al [9] found that a quantitative assessment (Method-2) of the ILD revealed a stronger correlation than a semiquantitative analysis for the assessment of therapeutic response. In addition to this finding, we found that the quantitative assessment using Method-2 showed a significant correlation with PFT results and semiquantitative disease extent scores. Ninaber and colleagues [8] also found that a quantitative assessment (Method-3) demonstrated a moderate to good correlation with PFT results. However, because they did not
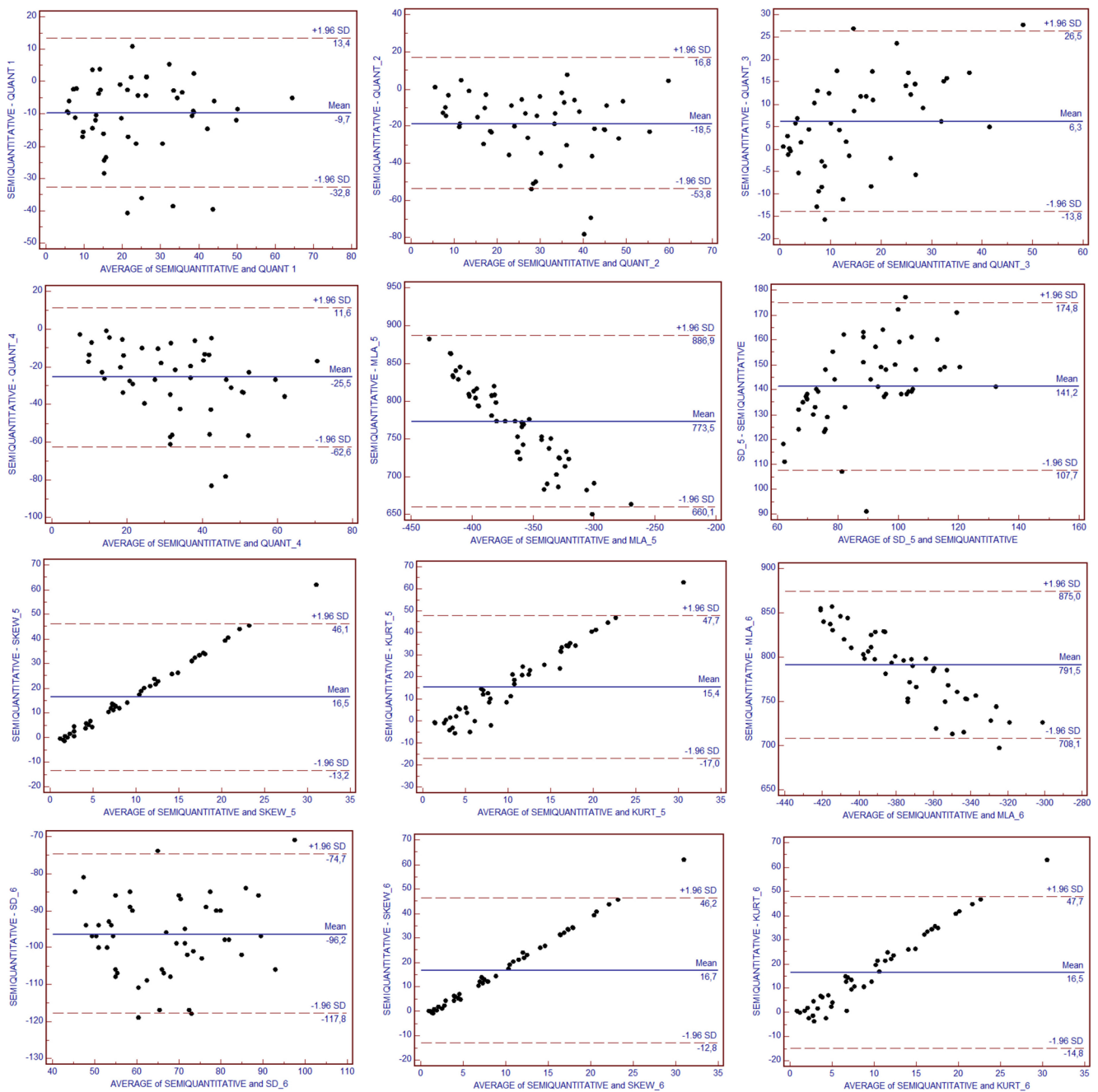
Fig. 4 Bland-Altman comparison of quantitative interstitial lung disease scores, pulmonary function tests, and visual scores for patients with systemic sclerosis–related interstitial lung disease. The average difference between the semiquantitative assessment and quantitative methods is also an estimate of agreemen

use a semiquantitative ILD assessment in their study [8], the relationship between the quantitative (Method-3) and semiquantitative assessment was unknown. We found a good correlation between Method-3 and a semiquantitative ILD assessment ($p < 0.0001$, $r = 0.619$). Furthermore, Marten and colleagues [12, 13] showed a significant and good correlation between a quantitative (Method-4) and semiquantitative ILD assessment ($p = 0.002$, $r = 0.65$ and $p < 0.0001$, $r = 0.716$, respectively). Similarly, we found a significant correlation

between Method-4 and the semiquantitative disease extent score ($p < 0.0001$, $r = 0.462$).

Koyama and colleagues [18] investigated quantitative ILD assessments using MLA, kurtosis, and skewness values in CT histogram analyses (Method-5) in 25 patients with CVD, and they found that quantitative values (MLA, kurtosis, skewness) revealed a fair to moderate correlation with PFT results. Similarly, we found a fair to good correlation between quantitative values (MLA-5, SKEW-5, KURT-5) and PFTs, except
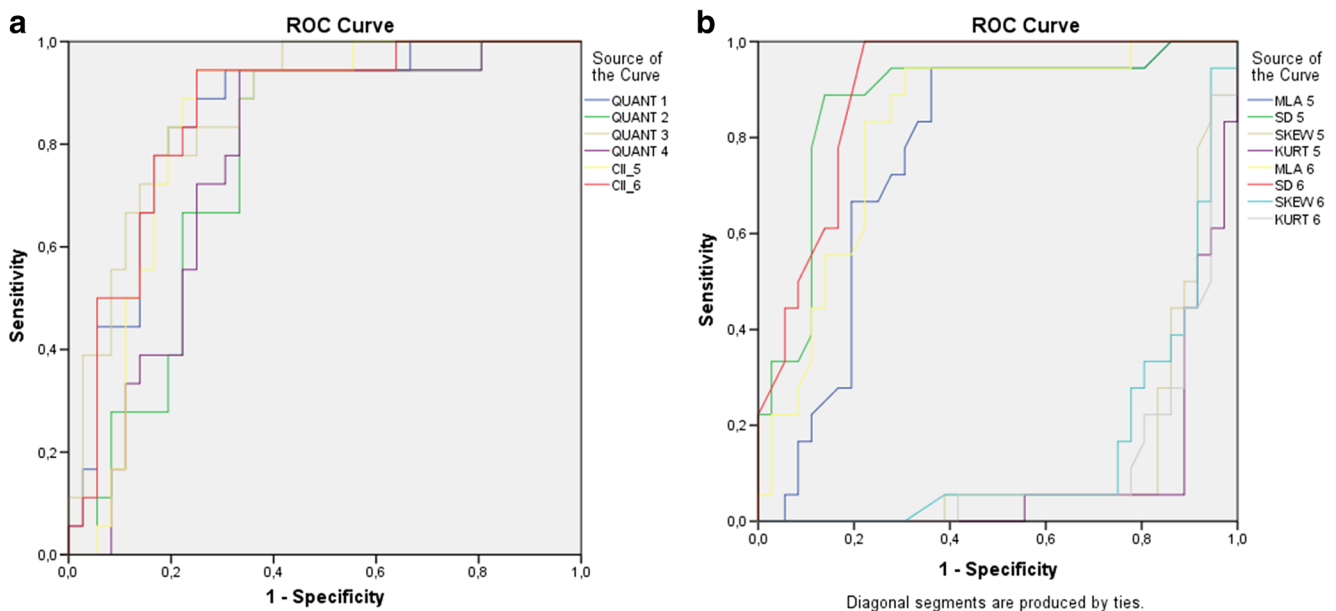
Fig. 5 Receiver operating characteristic curve testing the ability to differentiate the limited from extensive interstitial lung disease on high-resolution computed tomography. **a** Analysis for Method-1 to -4, computerized integrated index of Method-5, and computerized integrated index of Method-6. **b** Analysis for mean lung attenuation value of Method-5, mean lung attenuation value of Method-6, standard

deviation value of Method-5, standard deviation value of Method-6, kurtosis value of Method-5, kurtosis value of Method-6, skewness value of Method-5, and skewness value of Method-6. The ROC analysis demonstrated the best parameter was standard deviation value of Method-6 with an area under the curve (AUC) of 0.903 (95% CI, 0.824–0.982)

for the MLA-5 and FVC values. Because Koyoma et al did not use a semiquantitative ILD assessment in their study [18], the relationship between the quantitative (Method-5) and semiquantitative assessment was not investigated. In our study, we found a good to excellent correlation between quantitative (MLA-5, SKEW-5, KURT-5) and semiquantitative values. Ariani and colleagues [16] showed a significant correlation between the quantitative and semiquantitative ILD assessment, and they found the strongest degree of correlation with the semiquantitative disease extent score to be the skewness ($r = -0.378$) value of the voxels between -400 and -950 HU (SKEW-6). Although we found a good correlation between SKEW-6 and the semiquantitative ILD assessment score ($r = -0.643$), the kurtosis value of the voxels between -200 and -1024 HU (KURT-5) revealed an excellent correlation with the semiquantitative evaluation ($r = -0.740$).

A PFT also allows medical staff the ability to monitor patients with serial examination, which is important in terms of the prognosis and treatment of patients [31]. In PFTs, DLCO and FVC results are the most reported PFT parameters, which correlate with the quantitative and semiquantitative (visual) assessments [25]. The DLCO is considered to correlate the most with ILD severity on HRCT. However, the specificity of DLCO is low and has a high measurement error in the assessment of ILD severity. Pulmonary vascular disease or anemia has been shown to cause erroneous DLCO results [11, 25]. Moreover, it has been shown that FVC values in patients with CVDs might be influenced by muscle weakness and hardening or tightening of the thoracic skin [32].

Therefore, the severity of ILD in patients with SSc may sometimes not be associated with FVC values. Although, PFT is a noninvasive, usually safe and important tool for assessing the pulmonary condition in patients with CVD-related lung disease, performing PFTs may cause asthma attacks or dizziness, albeit rarely. Furthermore, elderly patients and children may not be able to perform the required maneuvers to meet recommendations, and performing PFTs in patients with a history of recent surgery and patients with congestive heart failure can be dangerous [28]. The results of QLA methods significantly correlate with PFT results in our study and QLA results provide information about PFT. Therefore, the need for PFT may decrease in patients (such as in patients with a history of recent surgery) undergoing HRCT to monitor disease progression and to assess prognosis.

Ariani et al [33] reported that QLA can differentiate mortality risk categories in patients with SSc. Furthermore, QLA is successful in the treatment follow-up in SSc patients and in the evaluation of pulmonary healing in SSc patients who underwent autologous stem cell transplantation [9, 20, 21]. Moreover, Bocchino et al [34] recently found that CII (a combined analysis of MLA, skewness, and kurtosis) values of QLA are significantly correlated with both PFTs and immune parameters (soluble cytokine receptors and C-C motif chemokine ligand 18), and they suggested that CII values could be sensitive for early detection of ILD in patients with SSc. Similarly, we found that CII-5 and CII-6 values revealed a significant correlation with other QLA methods, semiquantitative (visual) assessment scores, and PFTs (FVC, FEV1, and

Table 5 Area under the curve (AUC) values of receiver operating characteristic (ROC) curve for the discrimination of limited and extensive disease

Area under the curve

Variables	Area	SD error	p value	95% confidence interval	
				Lower bound	Upper bound
QUANT-1	0.853	0.054	0.0001	0.748	0.958
QUANT-2	0.758	0.067	0.002	0.626	0.889
QUANT-3	0.873	0.047	0.0001	0.782	0.965
QUANT-4	0.764	0.067	0.002	0.633	0.894
MLA-5	0.815	0.062	0.0001	0.694	0.936
SD-5	0.874	0.055	0.0001	0.766	0.982
SKEW-5	0.130	0.051	0.0001	0.030	0.231
KURT-5	0.088	0.041	0.0001	0.008	0.168
MLA-6	0.821	0.060	0.0001	0.704	0.938
SD-6	0.903	0.040	0.0001	0.824	0.982
SKEW-6	0.147	0.054	0.0001	0.042	0.253
KURT-6	0.119	0.048	0.0001	0.024	0.213
CII-5	0.861	0.052	0.0001	0.760	0.962
CII-6	0.863	0.052	0.0001	0.760	0.965

QUANT-1 quantitative Method-1, *QUANT-2* quantitative Method-2, *QUANT-3* quantitative Method-3, *QUANT-4* quantitative Method-4, *MLA-5* mean lung attenuation value of Method-5, *SD-5* standard deviation value of Method-5, *SKEW-5* skewness value of Method-5, *KURT-5* kurtosis value of Method-5, *MLA-6* mean lung attenuation value of Method-6, *SD-6* standard deviation value of Method-6, *SKEW-6* skewness value of Method-6, *KURT-6* kurtosis value of Method-6, *CII-5* computerized integrated index of Method-5, *CII-6* computerized integrated index of Method-6

DLCO). Although we found CII-5 and CII-6 values were highly discriminative for the limited and extensive ILD in patients with SSc, the SD-6 and SD-5 values were the best discriminative parameter. However, we suggest that SD values may be falsely high in patients with air trapping and mild ILD at the same time, since SD indicates heterogeneity (mosaic attenuation) in the lung parenchyma. In our study, Method-3 demonstrated the best and excellent performance for CT density cutoff methods and Method-3 was more successful than CII values for discrimination of limited and extensive ILD.

This study had several limitations. Although it is a retrospective study and the study population size was relatively small, this is the first study which compares QLA methods in patients with SSc-related ILD. Further prospective studies with a larger population are needed to substantiate our results. Another limitation of our study was the lack of follow-up and the effect of quantitative evaluation results on the prognosis of the patients. However, the effect of quantitative assessment results has been shown clearly in previous studies, and the disease extent on quantitative assessment proved to be a strong predictor of patient prognosis [21, 23–25, 32]. Besides, the lack of inter-

observer variability assessment is another limitation of our study. Lastly, we used a constant HRCT imaging technique (slice thicknesses and reconstruction parameters were constant). Therefore, we could not evaluate the effect of CT scanning and reconstruction parameters on the results, and future studies should evaluate the influence of the HRCT imaging technique and reconstruction parameters.

In conclusion, QLA (both histogram analysis and CT density cutoff) methods are applicable in assessing the ILD score in SSc patients and have the potential to differentiate limited from extensive ILD. Our results are helpful in the quantitative method selection for the assessment of ILD in SSc patients.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Furkan UFUK, MD.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry Dr. Hande Senol kindly provided statistical advice for this manuscript.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• Retrospective
• Cross sectional study
• Performed at one institution

References

1. Walker UA, Tyndall A, Czirjak L et al (2007) Clinical risk assessment of organ manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials and Research group database. Ann Rheum Dis 66:754–763
2. Steen VD, Medsger TA (2007) Changes in causes of death in systemic sclerosis, 1972-2002. Ann Rheum Dis 66:940–944
3. Ioannidis JP, Vlachoyiannopoulos PG, Haidich AB et al (2005) Mortality in systemic sclerosis: an international meta-analysis of individual patient data. Am J Med 118:2–10
4. Mouthon L, Berezné A, Brauner M, Kambouchner M, Guillevin L, Valeyre D (2007) Pneumopathie infiltrante diffuse de la sclerodermie systemique [Interstitial lung disease in systemic sclerosis]. Rev Mal Respir 24:1035–1046

5. Caron M, Hoa S, Hudson M, Schwartzman K, Steele R (2018) Pulmonary function tests as outcomes for systemic sclerosis interstitial lung disease. Eur Respir Rev 27:170102

6. Watadani T, Sakai F, Johkoh T et al (2013) Interobserver variability in the CT assessment of honeycombing in the lungs. Radiology 266:936–944

7. Aziz ZA, Wells AU, Hansell DM et al (2004) HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. Thorax 59:506–511

8. Ninaber MK, Stolk J, Smit J et al (2015) Lung structure and function relation in systemic sclerosis: application of lung densitometry. Eur J Radiol 84:975–979

9. Yabuuchi H, Matsuo Y, Tsukamoto H et al (2014) Evaluation of the extent of ground-glass opacity on high-resolution CT in patients with interstitial pneumonia associated with systemic sclerosis: comparison between quantitative and qualitative analysis. Clin Radiol 69:758–764

10. Salaffi F, Carotti M, Bosello S et al (2015) Computer-aided quantification of interstitial lung disease from high resolution computed tomography images in systemic sclerosis: correlation with visual reader-based score and physiologic tests. Biomed Res Int 2015: 834262

11. Salaffi F, Carotti M, Di Donato E, Di Carlo M, Ceccarelli L, Giuseppetti G (2016) Computer-aided tomographic analysis of interstitial lung disease (ILD) in patients with systemic sclerosis (SSc). Correlation with pulmonary physiologic tests and patient-centred measures of perceived dyspnea and functional disability. PLoS One 11:e0149240

12. Marten K, Dicken V, Kneitz C et al (2009) Interstitial lung disease associated with collagen vascular disorders: disease quantification using a computer-aided diagnosis tool. Eur Radiol 19:324–332

13. Marten K, Dicken V, Kneitz C et al (2009) Computer-assisted quantification of interstitial lung disease associated with rheumatoid arthritis: preliminary technical validation. Eur J Radiol 72:278–283

14. Çetinçakmak MG, Göya C, Hamidi C et al (2016) Quantitative volumetric assessment of pulmonary involvement in patients with systemic sclerosis. Quant Imaging Med Surg 6:50–56

15. Ariani A, Silva M, Bravi E et al (2015) Operator-independent quantitative chest computed tomography versus standard assessment of interstitial lung disease related to systemic sclerosis: a multi-centric study. Mod Rheumatol 25:724–730

16. Ariani A, Lumetti F, Silva M et al (2014) Systemic sclerosis interstitial lung disease evaluation: comparison between semiquantitative and quantitative computed tomography assessments. J Biol Regul Homeost Agents 28:507–513

17. Camiciottoli G, Orlandi I, Bartolucci M et al (2007) Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. Chest 131:672–681

18. Koyama H, Ohno Y, Yamazaki Y et al (2010) Quantitatively assessed CT imaging measures of pulmonary interstitial pneumonia: effects of reconstruction algorithms on histogram parameters. Eur J Radiol 74:142–146

19. Kazantzi A, Costaridou L, Skiadopoulos S et al (2014) Automated 3D interstitial lung disease extent quantification: performance evaluation and correlation to PFTs. J Digit Imaging 27:380–391

20. Kim HJ, Brown MS, Elashoff R et al (2011) Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. Eur Radiol 21:2455–2465

21. Kim HJ, Tashkin DP, Gjertson DW et al (2016) Transitions to different patterns of interstitial lung disease in scleroderma with and without treatment. Ann Rheum Dis 75:1367–1371

22. Kim HG, Tashkin DP, Clements PJ et al (2010) A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 28:S26–S35

23. Jacob J, Bartholmai BJ, Rajagopalan S et al (2016) Evaluation of computer-based computer tomography stratification against outcome models in connective tissue disease-related interstitial lung disease: a patient outcome study. BMC Med 14:190

24. Khanna D, Nagaraja V, Tseng CH et al (2015) Predictors of lung function decline in scleroderma-related interstitial lung disease based on high-resolution computed tomography: implications for cohort enrichment in systemic sclerosis-associated interstitial lung disease trials. Arthritis Res Ther 17:372

25. Tashkin DP, Volkmann ER, Tseng CH et al (2016) Relationship between quantitative radiographic assessments of interstitial lung disease and physiological and clinical features of systemic sclerosis. Ann Rheum Dis 75:374–381

26. Rosas IO, Yao J, Avila NA, Chow CK, Gahl WA, Gochuico BR (2011) Automated quantification of high-resolution CT scan findings in individuals at risk for pulmonary fibrosis. Chest 140:1590–1597

27. van den Hoogen F, Khanna D, Fransen J et al (2013) 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against rheumatism collaborative initiative. Ann Rheum Dis 72:1747–1755

28. Miller MR, Hankinson J, Brusasco V et al (2005) Standardisation of spirometry. Eur Respir J 26:319–338

29. Goh NS, Desai SR, Veeraraghavan S et al (2008) Interstitial lung disease in systemic sclerosis: a simple staging system. Am J Respir Crit Care Med 177:1248–1254

30. Zhang Z, Castelló A (2017) Principal components analysis in clinical studies. Ann Transl Med 5:351

31. Warrick JH, Bhalla M, Schabel SI, Silver RM (1991) High resolution computed tomography in early scleroderma lung disease. J Rheumatol 18:1520–1528

32. Winstone TA, Assayag D, Wilcox PG et al (2014) Predictors of mortality and progression in scleroderma-associated interstitial lung disease: a systematic review. Chest 146:422–436

33. Ariani A, Silva M, Seletti V et al (2017) Quantitative chest computed tomography is associated with two prediction models of mortality in interstitial lung disease related to systemic sclerosis. Rheumatology (Oxford) 56:922–927

34. Bocchino M, Bruzzese D, D'Alto M et al (2019) Performance of a new quantitative computed tomography index for interstitial lung disease assessment in systemic sclerosis. Sci Rep 9: 9468