Quantitative CT texture analysis for diagnosing systemic sclerosis

Effect of iterative reconstructions and radiation doses

Gianluca Milanese, MD,a,b Manoj Mannil, MD, MSc,a Katharina Martini, MD,a Britta Maurer, MD,c Hatem Alkadhi, MD,b Thomas Frauenfelder, MD*b

Abstract
To test whether texture analysis (TA) can discriminate between Systemic Sclerosis (SSc) and non-SSc patients in computed tomography (CT) with different radiation doses and reconstruction algorithms.

In this IRB-approved retrospective study, 85 CT scans at different radiation doses [49 standard dose CT (SDCT) with a volume CT dose index (CTDIvol) of 4.86 ± 2.1 mGy and 36 low-dose (LDCT) with a CTDIvol of 2.5 ± 1.5 mGy] were selected; 61 patients had SSc (“cases”), and 24 patients had no SSc (“controls”). CT scans were reconstructed with filtered-back projection (FBP) and with sinogram-verified iterative reconstruction (SAFIRE) algorithms. 304 TA features were extracted from each manually drawn region-of-interest at 6 predefined levels: at the midpoint between lung apices and tracheal carina, at the level of the tracheal carina, and 4 between the carina and pleural recesses. Each TA feature was averaged between these 6 pre-defined levels and was used as input in the machine learning algorithm artificial neural network (ANN) with backpropagation (MultilayerPerceptron) for differentiating between SSc and non-SSc patients.

Results were compared regarding correctly/incorrectly classified instances and ROC-AUCs.
ANN correctly classified individuals in 93.8% (AUC = 0.981) of FBP-LDCT, in 78.5% (AUC = 0.859) of FBP-SDCT, in 91.1% (AUC = 0.922) of SAFIRE3-LDCT and 75.7% (AUC = 0.815) of SAFIRE3-SDCT, in 88.1% (AUC = 0.929) of SAFIRE5-LDCT and 74% (AUC = 0.815) of SAFIRE5-SDCT.

Quantitative TA-based discrimination of CT of SSc patients is possible showing highest discriminatory power in FBP-LDCT images.

Abbreviations: GLCM = grey-level co-occurrence matrix, RLM = run-length matrix.

Keywords: neural networks (computer), systemic scleroderma, tomography, X-ray computed tomography (CT) with different radiation doses and reconstruction algorithms.

1. Introduction
Systemic sclerosis (SSc) is a connective-tissue disease characterized by excessive collagen production, fibrosis, and immunological abnormalities, frequently showing lung involvement.[1] Notably, interstitial lung disease (ILD) is the leading cause of death in patients with SSc.[1] Early detection of lung involvement is granted by high resolution computed tomography (CT), allowing characterization and quantification of ground glass opacities (GGO), reticulations, traction bronchiectasis, and microcystic honeycombing.[2] SSc patients are frequently screened with CT to assess presence and extent of ILD, in order to predict outcome and evaluate the need for immunosuppressive therapy.[3] Hence, optimizing and reducing cumulative radiation burden is relevant.[4]

Different approaches have been proposed for lowering the radiation burden, such as low-dose CT (LDCT) protocols and application of iterative reconstruction (IR) algorithms.[5,6] Previous studies indicated that the combination of LDCT protocols with IR results in similar detection rates for SSc-ILD as compared to the conventional reconstruction technique filtered-back projection (FBP) and standard dose CT (SDCT).[7]

CT has become the standard modality for evaluating lung parenchyma, nonetheless, evaluation and quantification of SSc-ILD with CT suffer from a relatively high inter-observer variability.[8,9] To overcome subjective visual assessment of CT images, densitometric analyses were shown to correlate with therapeutic response outperforming qualitative analyses.[9] This underlines the necessity to obtain an objective and quantitative method that allows for a reproducible evaluation of CT images in SSc-patients, ideally at a reduced radiation dose.

Recent interest was directed towards the quantitative evaluation of radiological images with texture analysis (TA), which is an objective approach for quantifying features being potentially...
imperceptible to the human eye and independent from the radiologists’ subjective evaluation. TA has been successfully applied in different medical applications, for example, to differentiate between benign and malignant pulmonary neoplasms and to predict outcome. The texture in images represents the spatial variation of pixel intensity values, which reflects tissural microstructure. TA quantifies the distribution of gray-level intensities within images as follows:

(a) based on histograms for pixel-values,
(b) evaluating the spatial variation of pixel values within a certain region-of-interest (ROI),
(c) based on the number of adjacent pixels showing the same pixel value, and
(d) based on the distribution of pairs of pixels.

Noteworthy, differences regarding image acquisition and image reconstruction parameters have been considered as possible causes of alterations not secondary to underlying tissural modifications.

The purpose of our study was 2-folded. First, we aimed to evaluate whether TA—performed on a small number of CT images—can differentiate between subjects with and without SSc independently from different image acquisition and image reconstruction parameters. Second, we tested the impact of different reconstruction algorithms and radiation doses on TA-results.

2. Materials and methods

2.1. Study population

In this retrospective study we included 85 patients (61 females, age range 18–80 years, mean 52.3 ± 14.3 years) referred to our department for single-energy non-enhanced chest CT between January 2012 and January 2013.

Sixty-one (71.8%) patients (hereinafter termed “cases”) had confirmed diagnosis of SSc according to the very early diagnosis of SSc (VEDOSS) and/or the American College of Rheumatology (ACR) classification criteria. Twenty-four (28.2%) patients (hereinafter termed “controls”) were referred to the radiology department for various other indications. Clinical indications for chest CT in the control group are listed in Table 1. Clinical indications for CT in controls are listed in Table 1.

Institutional review board approval was obtained.

2.2. CT scanning parameters and data reconstruction

CT scans were performed on a single-source 64-slice (Somatom Definition AS; Siemens Healthcare, Forchheim, Germany) and on a second-generation dual-source 128-slice CT scanner (Somatom Definition Flash; Siemens Healthcare) with automatic exposure control (CareDose4D, Siemens Healthcare). Tube current was modulated based on patients’ size. All scans were performed at end-inspiration in craniocaudal direction, capturing the entire lung volume from the apices to the pleural recesses.

CT datasets were reconstructed with FBP and with sinogram-affirmed IR (SAFIRE) algorithms; the latter comes with 5 pre-sets (strength levels) for predefined noise reduction. In our study, we included SAFIRE strength levels 3 and 5. SAFIRE allows for the evaluation of SSc with a similar accuracy as compared to FBP, and with SAFIRE strength level 5 the highest degree of image noise reduction can be obtained. The latter was included to test whether such high degree of noise reduction could be used in the setting of a TA-based diagnosis. All datasets were reconstructed using a sharp tissue convolution kernel (B60f for FBP and I70f for SAFIRE) with lung window settings (window width, 1200 HU; window level, −600 HU).

CT scanning and image reconstruction parameters, as well as radiation doses, are listed in the supplementary Table 1, http://links.lww.com/MD/D103, reporting the scanning protocols for both single-source CT (SS-CT) and second-generation dual-source CT scanner (DS-CT).

For each CT dataset, 6 slices at pre-defined levels, namely 1 at the midpoint between apices and the tracheal carina, 1 at the tracheal carina and 4 equally spaced between the tracheal carina and the pleural recesses, were selected and included for the TA. The evaluation of pulmonary parenchyma at different anatomical levels has been previously evaluated for various diseases.

2.3. Image analysis

A preliminary qualitative CT evaluation regarding the presence of GGO—the latter being a frequent CT finding in SSc—was performed by an expert radiologist (TF) on SDCT reconstructed with FBP.

TA was performed using a freely available software (MaZda, version 4.6, Institute of Electronics, Technical University of Lodz, Lodz, Poland), which allows for the two-dimensional segmentation of ROIs. Gray level normalization to correct for intra- and inter-scanner variations was performed between the mean and 3 standard deviations (“ ± 3σ” method). A board-certified radiologist (GM) performed the TA. On each of the pre-selected 6 slices, 1 bidimensional ROI per lung was manually drawn contouring the actual margin of the chest wall, to include the maximum amount of the sub-pleural space. The hilar vessels were carefully excluded from the ROI (Fig. 1). The ROIs were drawn on 1 reconstruction CT dataset and copy-pasted into the remaining 2 datasets. Texture results from the ROIs of the 6 CT slices were averaged for each subject.

Overall, 304 TA features were computed. TA features (supplementary Table 2, http://links.lww.com/MD/D103, detailing texture categories with corresponding features) originated from 6 main categories:

(a) Histogram,
(b) Grey-Level Co-Occurrence Matrix (GLCM) at 5 interpixel distances,
(c) Run-Length Matrix (RLM) at 4 angles: 0°, 90°, 135°, and 180°,
(d) Absolute gradient,
(e) Autoregressive model, and
(f) Wavelet transform.

---

**Table 1**

| Clinical indication for chest CT in the control group. | Number of subjects |
|-------------------------------------------------------|---------------------|
| Respiratory Bronchiolitis (RB)/ RB-ILD                | 5 (20.8%)            |
| ILD with UP pattern                                    | 1 (4.2%)             |
| COPD                                                   | 3 (12.5%)            |
| ILD associated with systemic lupus erythematous        | 1 (4.2%)             |
| ILD associated with connective tissue disease         | 1 (4.2%)             |
| Cystic fibrosis                                        | 1 (4.2%)             |
| Pneumonia                                              | 4 (16.6%)            |
| Follow-up of oncologic disease                         | 2 (8.3%)             |
| LAM                                                    | 1 (4.2%)             |
| Follow-up after lung transplant                        | 4 (16.6%)            |
| Asbestosis                                             | 1 (4.2%)             |

* One patient suffered from b-cell lymphoma, 1 had history of breast cancer.

COPD = chronic obstructive pulmonary disease, LAM = lymphangioleiomyomatosis, RB-ILD = respiratory bronchiolitis—interstitial lung disease, UP = usual interstitial pneumonia.
Dimension reduction was performed using the supervised class filter “CfsSubsetEval” of the open source software Weka (University of Waikato, Waikato, New Zealand). It evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Feature selection is performed on the training data set by assessing subset of features that are highly correlated with the class while having low intercorrelation (Supplementary Figure 1, http://links.lww.com/MD/D103).

### 2.4. TA feature reduction

Dimension reduction was performed using the supervised class filter “CfsSubsetEval” of the open source software Weka (University of Waikato, Waikato, New Zealand). It evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Feature selection is performed on the training data set by assessing subset of features that are highly correlated with the class while having low intercorrelation (Supplementary Figure 1, http://links.lww.com/MD/D103).

### 2.5. Statistical analysis

Normality of data distribution was assessed with the Shapiro-Wilk test. Normally distributed variables were reported as mean and standard deviation (SD), non-normally distributed variables were reported as median and inter-quartile range (IQR). The variables were compared applying Student t-test for parametric variables and Mann–Whitney U test for non-parametric variables. Categorical variables were compared using Chi-squared test.

After feature reduction, all the remaining TA features acted as input in an artificial neural network (ANN, “Multilayer Perceptron”, MLP) with backpropagation for classification of cases and controls. MLP is a traditional neural network characterized by different layers inputs and output and is currently the most well-known among traditional neural networks.\(^{(26)}\)

Results of the ANN-based classifications were compared regarding correctly/incorrectly classified instances, mean absolute error, true positive (TP), and false positive (FP) rates, precision and the receiver-operating characteristics—area under the curve (ROC-AUC). The whole dataset was initially used for training and testing. Then, it was split in the recommended ratio of 2/3 training and 1/3 testing set. Subsequently, to account for overfitting, classification analyses were performed, and the results were compared after the 10-fold cross-validation process, which allows for all observations to serve for both training and validation, and each observation, is used for validation exactly once. We used stratified 10-fold cross-validation, in which the folds are selected so that each fold contains roughly the same proportions of class labels.\(^{(27)}\)

The results are averages of 10 runs. This analysis resulted in Model 1. Afterward, the analyses were repeated for the whole study population divided according to the reconstruction algorithm (Model 2) and between CT datasets with different radiation doses (i.e., LDCT and SDCT) and reconstruction algorithms (Model 3).

Data mining and ANN analyses were performed using open-source software (WEKA, University of Waikato), all remaining statistical analyses were conducted using commercially available software (SPSS 23.0; IBM, Chicago, Ill). A 2-tailed P value below .05 was considered to indicate statistical significance.

### 3. Results

In this retrospective study we included 85 patients (61 females, age range 18–80 years, mean 52.3±14.3 years) that underwent CT scans at different radiation doses [49 SDCT with a volume CT dose index (CTDIvol) of 4.86±2.1 mGy and 36 LDCT with a CTDIvol of 2.5±1.5 mGy]. There were no significant differences in mean age between men and women (\(P=.29\)) and between cases and controls (\(P=.12\)).

Among cases, 33 (54.1%) had limited SSc and 28 (45.9%) patients had diffuse SSc based on the LeRoy criteria \(^{(28)}\). Furthermore, 28 individuals reported symptoms from esophageal insufficiency and 48 (78.7%) reported dyspnea.

#### 3.1. Image analysis

Qualitative evaluation showed that 41 of the 61 cases with confirmed SSc had no GGO (SDCT: 16/41 and LDCT: 25/41).

Twenty cases (SDCT: 9/20 and LDCT: 11/20) and 12 controls (SDCT: 12/12) showed GGO.

TA could be successfully performed in all CT datasets.

#### 3.2. Differences in discriminatory power of TA (Model 1)

For the whole dataset, 13 out of the initial 304 (4.3%) TA features were included in our analysis after feature reduction. By applying 10-fold cross-validation classification the AUC was 0.878, with a correct classification in 85% of cases.

#### 3.3. Differences in discriminatory power of TA using different CT reconstruction filters (Model 2)

For FBP, 17 out of the initial 304 (5.6%) TA features were included in the analysis after feature selection; for IR, 13 TA features (4.3%) were included in the analysis for both SAFIRE 3 and SAFIRE 5 (supplementary Table 3, http://links.lww.com/MD/D103), reporting the results of dimension reduction by reconstruction type.

The results of Model 2 are listed in Table 2. The highest discriminatory power for TA-based diagnosis of SSc was possible with FBP. With 10-fold cross-validation classification, the highest AUC among CT datasets was found for the FBP-reconstructed images (0.904) with a correct classification in the 84.9% of cases (Fig. 2). For SAFIRE 3 datasets, the AUC was 0.851 and 83.3% of cases were correctly classified. For SAFIRE 5 datasets, the AUC was 0.799 and 78.4% of cases were correctly classified.
3.4. Differences in discriminatory power of TA between SDCT and LDCT (Model 3)

For FBP, 14 TA features with SDCT and 22 TA features with LDCT were included into the analysis after feature selection. For SAFIRE 3, 8 features for SDCT and 10 for LDCT were included. For SAFIRE 5, 11 features for both SDCT and LDCT were included.

The results of Model 3 are listed in Table 3. The highest discriminatory power for TA-based diagnosis of cases was found for FBP in comparison with SAFIRE 3 and SAFIRE 5, on both SDCT and LDCT datasets.

4. Discussion

Our study shows that the TA-based diagnosis of SSc is feasible and can be performed on a small number of CT slices which is beneficial for CT scanning at low radiation dose. Then, the discriminatory power for differentiating patients with SSc and those without was higher for LDCT, showing a progressive reduction in accuracy from FBP to SAFIRE 3 and SAFIRE 5.
One major field of interest of ANN analysis in imaging is to evaluate whether it can discriminate between the presence and absence of a certain disease based on input from images, especially regarding textural patterns.\textsuperscript{12,26,29} In our study, “handcrafted” TA (i.e., ROIs were manually drawn) could discriminate between cases and controls based on texture features extracted from 6 CT images at different anatomical levels. This spatial sampling of texture is relevant for the assessment of diffuse lung diseases.\textsuperscript{30}

In daily clinical setting, the diagnosis of SSc is based on clinical, serological, functional and imaging results.\textsuperscript{31} Furthermore, parenchymal abnormalities may not be detectable through alterations of pulmonary function test in patients with early SSc. Imaging can depict parenchymal changes suggestive for the development of ILD during follow-up studies; nevertheless, the radiation burden is of particular concern.\textsuperscript{31} In keeping with the literature, we report the possible clinical application of quantitative evaluation and texture- and density-based analysis. Moon et al reported that histograms of ILD patients were less skewed and less Kurtotic than those of non-ILD patients and they postulated that these differences related with an increased heterogeneous attenuation.\textsuperscript{31} Furthermore, Ariani et al demonstrated that quantitative CT indexes could stratify SSc patients according to their prognosis.\textsuperscript{32} A neural network proposed by Anthimopoulos et al was shown to possess the potential in standard-dose CT, TP\textsuperscript{33,34} was shown to be able to be aware that potential variations in extracted features during follow-up may not be related to actual changes in the course of the disease if different reconstruction settings were used. Kim et al showed significant differences among features extracted from the same patient on non-contrast and contrast-enhanced CT.\textsuperscript{37} Furthermore, Larue et al reported that features are influenced by CT-scanner, slice thickness, and bin width.\textsuperscript{38} Similarly, Mackin et al highlighted the relevance of image acquisition and reconstruction for the repeatability of features.\textsuperscript{11} A recent inter-CT and intra-CT study evaluating radiomic features showed that many features were not reproducible, and the authors underlined the importance of a standardized methodology, notably for CT reconstruction kernels and section thickness.\textsuperscript{11} Similarly, Ahn et al underlined the potential effect of reconstruction algorithms and CT parameters on texture analysis.\textsuperscript{40} Of note, in our study we included CT scans obtained from 2 different CT-scanners, nevertheless, the ability of the ANN analysis to classify cases and controls was correct in more than 90% of patients, and when analyzing the study population independently from the different image acquisition parameters, it was 85%. We underline the potential relevance of robust TA, as it could be particularly beneficial in multicentric settings, in which similar scanning parameters may not be easily obtained. To expand quantitative approaches allowing reproducible multicentric studies, a recent phantom study showed that a dedicated compensation algorithm realigning quantitative features can be used to remove the effects of scanning protocols.\textsuperscript{41} Ideally, the same protocol should be used at baseline and during quantitative follow-up. Quantitative features have been reported to vary according to image pre-processing and reconstruction techniques potentially introducing changes not caused by underlying biologic effects,\textsuperscript{11,42} and a study testing the effect of IR algorithms toward quantitative image features, showed that features derived from colorectal exams were altered according to IR levels.\textsuperscript{43} Similarly, Mannil et al reported the impact of reconstruction algorithm toward TA on myocardial infarction in cardiac CT.\textsuperscript{44}

In our study, a number of histogram-derived features were included in the ANN classifier, potentially representing a measurement of the parenchymal density (evaluated as Hounsfield Unit) and therefore being suggestive of GGO, which is a frequent finding in SSc-ILD patients.\textsuperscript{23} Notably, only 20 out of 61 SSc patients displayed GGO, as qualitatively assessed. On the other hand, the classifier included mostly “not histogram-
derived” features, depicting structural and textural changes that are—most probably—not perceptible to human eyes. Furthermore, reduction of noise by application of IR may be associated with a concomitant loss of information within images, potentially resulting in a higher performance of FBP. Hence, we postulate that TA could be influenced by the effect of iterative reconstruction—in terms of noise reduction—for quantitative analyses. In our study—performed in a clinical setting—the number of protocols was limited, and a preliminary gray-level normalization was performed. Larue et al recommend performing a preliminary “gray-level discretization” before the analysis. Peikert et al reported that features derived from a heterogeneous sample of the CT dataset obtained during the NLST were robust and stable across different scanners, image acquisition parameters, and reconstructions algorithms. Of note, different scanners, and acquisition parameters of magnetic resonance images of patients with glioblastoma leading to heterogeneous condition were considered as a possible strength, related to a realistic clinical condition—as in our study—given that in large multicentric databases the same imaging protocol can be difficult to be achieved.

In our study, the AUC of LDCT was greater than that of SDCT. Pontana et al reported a higher percentage of excellent visual scores for the assessment of interstitial anomalies in SSc patients evaluated on IR-reduced dose images, as compared to FBP-SDCT. Christe et al reported a reduced subjective detection rate for interstitial abnormalities on LDCT, in particular for reticulation (reduced sensitivity of 20%). Scanning patients with LDCT is not only beneficial for the purpose of reducing the radiation burden for individuals requiring long-term annual radiological follow-up, but it seems to increase the discriminatory power of TA. IR reduces image noise and the radiation burden for individuals requiring long-term annual scanning protocols with 120 kVp, thus allowing a decrease in the radiation dose of 30% to 40% compared with scanning protocols with 120 kVp. Third, as the ROIs were drawn by a single reader, we could not evaluate whether TA features would significantly differ when the ROIs are drawn by multiple readers. Fourth, CT datasets were reconstructed with a certain version of the IR algorithm, and other IR algorithms were not tested herein. Nevertheless, a recent study on abdominal CT did not detect significant differences regarding image noise, signal-to-noise ratio, and contrast-to-noise ratio. Fifth, we included in our analysis SAFIRE 5, which is not currently used in daily clinical practice mainly because of the “plastic” appearance associated with the higher levels of iterative reconstructions. Nevertheless, for the evaluation of the TA-based analysis, we included the highest level of IR that was available to evaluate whether it could be used in the specific setting of cases evaluated with a reduced number of CT images. Lastly, we did not test whether TA-based diagnosis could differentiate between SSc patients with ILD and cases without ILD. Our purpose was to evaluate whether TA could differentiate between cases irrespectively of the presence of parenchymal changes and to evaluate the best scanning protocols that allowed for the differentiation between cases and controls displaying various degrees of parenchymal abnormalities. We foster future studies for the evaluation of TA based diagnosis of parenchymal abnormalities in patients with SSc and other ILDs.

In conclusion, our study indicates that objective and quantitative CT-based TA allows for the accurate diagnosis of patients with SSc, with best results for LDCT and using the FBP reconstruction algorithm.

Author contributions
Conceptualization: Thomas Frauenfelder.
Data curation: Manoj Mannil.
Formal analysis: Manoj Mannil, Thomas Frauenfelder.
Investigation: Gianluca Milanese, Manoj Mannil, Thomas Frauenfelder.
Methodology: Gianluca Milanese, Hatem Alkadhi, Thomas Frauenfelder.
Supervision: Hatem Alkadhi, Thomas Frauenfelder.
Writing – original draft: Gianluca Milanese.
Writing – review & editing: Manoj Mannil, Katharina Martini, Britta Maurer, Hatem Alkadhi, Thomas Frauenfelder.
Gianluca Milanese orcid: 0000-0003-1974-4854.

References
[1] Solomon JJ, Olson AL, Fischer A, et al. Scleroderma lung disease. Eur Respir Rev 2013;22:6–19.
[2] Bussone G, Mouthon L. Interstitial lung disease in systemic sclerosis. Autoimmun Rev 2011;10:248–55.
[50] Geyer LL, Schoepf UJ, Meinel FG, et al. State of the art: iterative CT reconstruction techniques. Radiology 2015;276:339–57.
[51] Pavarani A, Martini C, Gafa V, et al. Effect of iterative reconstruction on image quality of low-dose chest computed tomography. Acta Biomed 2016;87:168–76.
[52] Mets OM, Willemink MJ, de Kort FP, et al. The effect of iterative reconstruction on computed tomography assessment of emphysema, air trapping and airway dimensions. Eur Radiol 2012;22:2103–9.
[53] Nguyen-Kim TDL, Maurer B, Suliman YA, et al. The impact of slice-reduced computed tomography on histogram-based densitometry assessment of lung fibrosis in patients with systemic sclerosis. J Thorac Dis 2018;10:2142–52.
[54] Litmanovich DE, Tack DM, Shahrzad M, et al. Dose reduction in cardiothoracic CT: review of currently available methods. Radiographics 2014;34:1469–89.