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[a], Thomas Frauenfelder, MD[a,*]

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-affirmed iterative reconstruction (SAFIRE) algorithms. 304 TA features were extracted from each manually drawn region-of-interest at 6 pre-defined 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

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.[10–12] TA has been successfully applied in different medical applications, for example, to differentiate between benign and malignant pulmonary neoplasms and to predict outcome.[13,14] The texture in images represents the spatial variation of pixel intensity values, which reflects tissural microstructure.[10] 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.[15]

Noteworthy, differences regarding image acquisition and image reconstruction parameters have been considered as possible causes of alterations not secondary to underlying tissural modifications.[16,17]

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.[18,19] Twenty-four (28.2%) patients (hereinafter termed "controls") were referred to the radiology department for various other indications. 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 studies were performed on a single-source 64-slice (Somatom Definition AS; Siemens Healthcare, Forchheim, Germany) and on

Table 1

Clinical indication for chest CT in the control group.

Clinical indications for chest CT	Number of subjects
Respiratory Bronchiolitis (RB)/RB–ILD	5 (20.8%)
ILD with UIP pattern	1 (4.2%)
COPD	3 (12.5%)
ILD associated with systemic lupus erythematosus	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, UIP = usual interstitial pneumonia.

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 (15). 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.[7,20] 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 datasets, 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.[21,22]

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.[23]

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.[24] Gray level normalization to correct for intra- and inter-scanner variations was performed between the mean and 3 standard deviations ("$\pm 3\sigma$" method).[25] 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.

Figure 1. Axial chest CT image in prone position. The 2 green areas comprising the lung parenchyma represent the ROIs used for texture analysis at the level of the tracheal carina. CT = computed tomography, ROI = region of interest.

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 CT 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 FPB. 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.

Table 2

Results of the TA analysis for CT datasets reconstructed with FBP, SAFIRE 3, and SAFIRE 5 for the whole study population.

		Entire dataset					Percentage split (ratio $\frac{2}{3}$ and $\frac{1}{3}$)					10-fold cross-validation				
Reconstruction algorithm	Subjects	Correctly classified (%)	TP rate	FP rate	Precision	AUC (95%CI)	Correctly classified (%)	TP rate	FP rate	Precision	AUC (95%CI)	Correctly classified (%)	TP rate	FP rate	Precision	AUC (95%CI)
FBP	Case (n=61)	95.3	0.967	0.083	0.967	0.967 (0.85–1)	85.5	0.925	0.297	0.887	0.870 (0.76–0.98)	84.9	0.910	0.306	0.883	0.904 (0.79–1)
	Control (n=24)		0.917	0.033	0.917	0.967 (0.85–1)		0.727	0.114	0.600	0.916 (0.8–1)		0.694	0.090	0.752	0.904 (0.79–1)
SAFIRE 3	Case (n=61)	95.1	0.984	0.132	0.950	0.954 (0.84–1)	81.5	0.864	0.394	0.903	0.828 (0.71–0.93)	83.3	0.899	0.333	0.873	0.851 (0.74–0.96)
	Control (n=24)		0.868	0.016	0.917	0.954 (0.84–1)		0.606	0.136	0.513	0.828 (0.71–0.93)		0.667	0.101	0.722	0.851 (0.74–0.96)
SAFIRE 5	Case (n=61)	90.4	0.970	0.264	0.903	0.891 (0.78–1)	83.8	0.886	0.364	0.912	0.848 (0.73–0.95)	78.4	0.861	0.410	0.842	0.799 (0.68–0.91)
	Control (n=24)		0.736	0.030	0.906	0.891 (0.78–1)		0.636	0.114	0.568	0.848 (0.73–0.95)		0.590	0.139	0.625	0.799 (0.68–0.91)

AUC = area under the curve from receiver operating characteristics analysis, FBP = filtered-back projection, FP = false positive, SAFIRE = Sinogram Affirmed Iterative Reconstruction, TP = true positive.

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 FPB 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 FPB to SAFIRE 3 and SAFIRE 5.
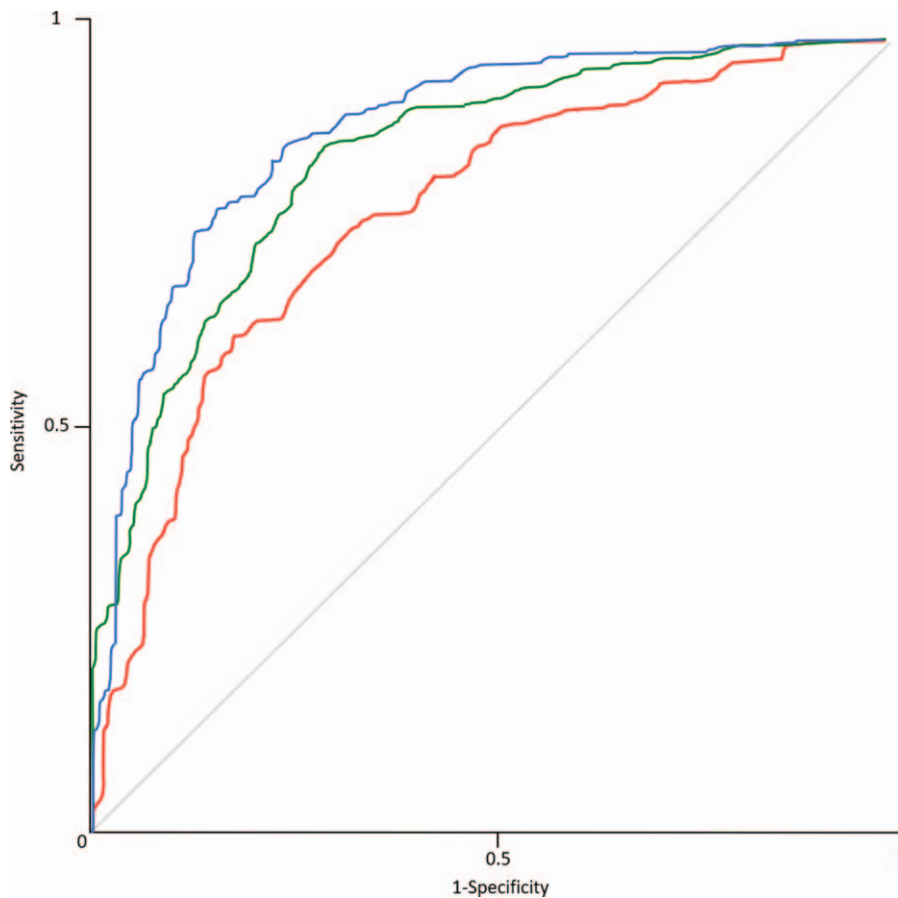
Figure 2. ROC analysis curves of artificial neural network with 10-fold cross-validation averaged for standard and low dose CT by reconstruction type: Filtered Back Projection (Blue), SAFIRE strength level 3 (Green), and SAFIRE strength level 5 (Red). Note the highest AUC at FBP reconstructed images. AUC = area under the curve, CT = computed tomography, ROC = receiver operating characteristics, SAFIRE = Sinogram Affirmed Iterative Reconstruction.

Table 3

Results of the TA analysis for CT datasets reconstructed with FBP, SAFIRE 3 and SAFIRE 5 stratified according to the radiation dose of the CT.

Reconstruction algorithm	Subjects	Entire dataset					Percentage split (ratio $\frac{2}{3}$ and $\frac{1}{3}$)					10-fold cross-validation				
		Correctly classified (%)	TP rate	FP rate	Precision	AUC (95%CI)	Correctly classified (%)	TP rate	FP rate	Precision	AUC (95%CI)	Correctly classified (%)	TP rate	FP rate	Precision	AUC (95%CI)
FBP SDCT	Case (n=61)	94.1	0.917	0.035	0.964	0.959 (0.84–1)	74.5	0.783	0.288	2.706	0.801 (0.69–0.91)	78.5	0.764	0.194	0.797	0.859 (0.73–0.96)
	Control (n=24)		0.965	0.083	0.921	0.959 (0.84–1)		0.712	0.217	0.787	0.801 (0.69–0.91)		0.806	0.236	0.773	0.859 (0.73–0.96)
FBP LDCT	Case (n=61)	99.4	1	0.014	0.991	0.992 (0.89–1)	92.5	0.915	0.061	0.959	0.980 (0.87–1)	93.8	0.953	0.083	0.943	0.981 (0.87–1)
	Control (n=24)		0.986	0	1	0.992 (0.89–1)		0.939	0.085	0.885	0.980 (0.87–1)		0.917	0.048	0.930	0.981 (0.87–1)
SAFIRE 3 SDCT	Case (n=61)	85.1	0.840	0.139	0.858	0.910 (0.81–1)	68.4	0.609	0.250	0.683	0.745 (0.63–0.85)	75.7	0.743	0.229	0.764	0.815 (0.7–0.92)
	Control (n=24)		0.861	0.160	0.844	0.910 (0.81–1)		0.750	0.391	0.684	0.745 (0.63–0.85)		0.771	0.257	0.750	0.815 (0.7–0.92)
SAFIRE 3 LDCT	Case (n=61)	96.7	0.986	0.063	0.959	0.993 (0.88–1)	87.7	0.961	0.267	0.860	0.915 (0.8–1)	91.1	0.958	0.160	0.900	0.922 (0.8–1)
	Control (n=24)		0.938	0.014	0.978	0.993 (0.88–1)		0.733	0.039	0.917	0.915 (0.8–1)		0.840	0.042	0.931	0.922 (0.8–1)
SAFIRE 5 SDCT	Case (n=61)	87.5	0.813	0.063	0.929	0.953 (0.84–1)	67.3	0.761	0.404	0.625	0.741 (0.63–0.85)	74.0	0.708	0.229	0.756	0.815 (0.7–0.9)
	Control (n=24)		0.938	0.188	0.833	0.953 (0.84–1)		0.596	0.239	0.738	0.741 (0.63–0.85)		0.771	0.292	0.725	0.81 (0.7–0.92)
SAFIRE 5 LDCT	Case (n=61)	96.9	1	0.076	0.952	0.969 (0.85–1)	90.2	0.935	0.156	0.965	0.959 (0.84–1)	88.1	0.921	0.181	0.884	0.929 (0.81–1)
	Control (n=24)		0.924	0	1	0.969 (0.85–1)		0.965	0.083	0.921	0.959 (0.84–1)		0.819	0.079	0.874	0.929 (0.91–1)

AUC = area under the curve from receiver operating characteristics analysis, FBP = filtered-back projection, FP = false positive, LDCT = low-dose, SAFIRE = Sinogram Affirmed Iterative Reconstruction, SDCT = standard-dose CT, TP = true positive.

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.[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.[30]

In daily clinical setting, the diagnosis of SSc is based on clinical, serological, functional and imaging results.[3] 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.[3] 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.[31] Furthermore, Ariani et al demonstrated that quantitative CT indexes could stratify SSc patients according to their prognosis.[32] A neural network proposed by Anthimopoulos et al was shown to possess the potential in classifying ILD patterns, including GGO and reticulations.[33,34] Best et al showed that changes in density-based quantitative CT indices such as mean lung attenuation, skewness, and kurtosis were associated with the progression of idiopathic pulmonary fibrosis.[35] Recently, classification of ILD based on a deep learning approach—obtaining a human level performance—was reported by Walsh et al.[34]

Of note, the classification performed by the ANN of the whole study population—distinguishing between cases and controls—was correct in 85% of cases, independently from radiation doses and reconstruction algorithms. Such results underline the robustness and reproducibility of TA. Nonetheless, our study shows slight differences in the AUC when stratifying between CT datasets with different radiation doses. Solomon et al stated that reconstruction parameters should not be changed between studies as different radiation doses and reconstruction algorithms affect TA.[36] The authors underlined the importance for radiologists 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.[37] Furthermore, Larue et al reported that features are influenced by CT-scanner, slice thickness, and bin width.[38] Similarly, Mackin et al highlighted the relevance of image acquisition and reconstruction for the repeatability of features.[16] 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.[39] Similarly, Ahn et al underlined the potential effect of reconstruction algorithms and CT parameters on texture analysis.[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.[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,[17,42] and a study testing the effect of IR algorithms toward quantitative image features, showed that features derived from colorectal cancers were altered according to IR levels.[43] Similarly, Mannil et al reported the impact of reconstruction algorithm toward TA on myocardial infarction in cardiac CT.[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.[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.[27] Larue et al recommend performing a preliminary "gray-level discretization" before the analysis.[38] 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.[45] 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.[46,47,48]

In our study, the AUC of LDCT was greater than that of SDCT. Pontana et al reported a higher frequency of excellent visual scores for the assessment of interstitial anomalies in SSc patients evaluated on IR-reduced dose images, as compared to FBP-SDCT.[7] Christe et al reported a reduced subjective detection rate for interstitial abnormalities on LDCT, in particular for reticulation (reduced sensitivity of 20%).[49] 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 [5,50]; however, our results show the highest ROC-AUC for FBP-LDCT on a reduced (i.e. 6) number of slices, with pulmonary parenchyma sampled at pre-defined intervals. Of note, appearance of pulmonary parenchyma was reported to be affected by IR particularly in the setting of the quantitative evaluation of pulmonary emphysema as assessed by low-attenuation areas.[51,52]

A reduced slice-approach was tested by Winklehner et al, reporting the viability of the detection of ILD on CT image series with low sampling rate.[21] Our results expand those of Nguyen-Kim et al, who recently reported that both visual and histogram-based assessment is possible on slice-reduced sequential CT. Histogram parameters such as kurtosis and skewness were capable of discriminating fibrosis (20% cut-off) with high sensitivity and specificity.[53] We included in our analysis the highest strength level for IR, namely SAFIRE 5, reported to provide a so-called "blotchy-appearance", which could hamper not only visual but also quantitative evaluation of pulmonary abnormalities. We postulate that by applying newer generation of IR (such as model-based IR), effect of TA could be improved.[21,53]

Importantly, CT scans of controls were requested for different clinical indications and displayed parenchymal anomalies, such as GGO and reticulations being abnormalities that can be seen also in patients with SSc-ILD.[2] Hence, controls did not properly consist of a "healthy" population, but TA could discriminate them from cases. In our study, controls underwent CT scanning because of various indications, such as RB-ILD which may manifest as diffuse GGO [47] and obliterative bronchiolitis (OB) —a possible complication after lung transplant—detectable as a mosaic attenuation pattern.[48] Nevertheless, TA could discriminate between cases and controls, even in this daily clinical setting with parenchymal anomalies detectable in both groups.

Our study has various limitations. First, in the analysis performed to discriminate the TA-based diagnosis of SSc according to the CT dose levels, machine-learning based discrimination between cases and controls may mirror inherent differences between SDCT and LDCT, irrespective of the presence of SSc. Second, controls were scanned with tube current-time product higher than 40 mAs; however, the kilovoltage was set—for controls evaluated with both scanners —at 100 kVp, thus allowing a decrease in the radiation dose of 30% to 40% compared with scanning protocols with 120 kVp.[54] 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.[20] 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.[20] 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.

[3] Molberg O, Hoffmann-Vold AM. Interstitial lung disease in systemic sclerosis: progress in screening and early diagnosis. Curr Opin Rheumatol 2016;28:613–8.

[4] Frauenfelder T, Winklehner A, Nguyen TD, et al. Screening for interstitial lung disease in systemic sclerosis: performance of high-resolution CT with limited number of slices: a prospective study. Ann Rheum Dis 2014;73:2069–73.

[5] Baumueller S, Winklehner A, Karlo C, et al. Low-dose CT of the lung: potential value of iterative reconstructions. Eur Radiol 2012;22:2597–606.

[6] Milanese G, Silva M, Frauenfelder T, et al. Comparison of ultra-low dose chest CT scanning protocols for the detection of pulmonary nodules: a phantom study. Tumori 2019;May 1:300891619847271. doi: 10.1177/0300891619847271. [Epub ahead of print].

[7] Pontana F, Billard AS, Duhamel A, et al. Effect of iterative reconstruction on the detection of systemic sclerosis-related interstitial lung disease: clinical experience in 55 patients. Radiology 2016;279:297–305.

[8] Ariani A, Lumetti F, Silva M, et al. Systemic sclerosis interstitial lung disease evaluation: comparison between semiquantitative and quantitative computed tomography assessments. J Biol Regul Homeost Agents 2014;28:507–13.

[9] Yabuuchi H, Matsuo Y, Tsukamoto H, et al. 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 2014;69:758–64.

[10] Sogawa K, Nodera H, Takamatsu N, et al. Neurogenic and myogenic diseases: quantitative texture analysis of muscle us data for differentiation. Radiology 2017;283:492–8.

[11] Baessler B, Mannil M, Oebel S, et al. Subacute and chronic left ventricular myocardial scar: accuracy of texture analysis on non-enhanced cine MR images. Radiology 2018;286:103–12.

[12] Silva M, Milanese G, Seletti V, et al. Pulmonary quantitative CT imaging in focal and diffuse disease: current research and clinical applications. Br J Radiol 2018;91:20170644.

[13] Parekh V, Jacobs MA. Radiomics: a new application from established techniques. Expert Rev Precis Med Drug Dev 2016;1:207–26.

[14] Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res 2017;6:86–91.

[15] MacKay JW, Murray PJ, Low SB, et al. Quantitative analysis of tibial subchondral bone: texture analysis outperforms conventional trabecular microarchitecture analysis. J Magn Reson Imaging 2016;43:1159–70.

[16] Mackin D, Fave X, Zhang L, et al. Measuring computed tomography scanner variability of radiomics features. Invest Radiol 2015;50:757–65.

[17] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278:563–77.

[18] Minier T, Guiducci S, Bellando-Randone S, et al. Preliminary analysis of the very early diagnosis of systemic sclerosis (VEDOSS) EUSTAR multicentre study: evidence for puffy fingers as a pivotal sign for suspicion of systemic sclerosis. Ann Rheum Dis 2014;73:2087–93.

[19] Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23:581–90.

[20] Morsbach F, Desbiolles L, Raupach R, et al. Noise texture deviation: a measure for quantifying artifacts in computed tomography images with iterative reconstructions. Invest Radiol 2017;52:87–94.

[21] Winklehner A, Berger N, Maurer B, et al. Screening for interstitial lung disease in systemic sclerosis: the diagnostic accuracy of HRCT image series with high increment and reduced number of slices. Ann Rheum Dis 2012;71:549–52.

[22] Walsh SL, Sverzellati N, Devaraj A, et al. Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. Eur Radiol 2012;22:1672–9.

[23] Desai SR, Veeraraghavan S, Hansell DM, et al. CT features of lung disease in patients with systemic sclerosis: comparison with idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia. Radiology 2004;232:560–7.

[24] Szczypinski PM, Strzelecki M, Materka A, et al. MaZda–a software package for image texture analysis. Comput Methods Programs Biomed 2009;94:66–76.

[25] Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004;22:81–91.

[26] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.

[27] Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics 2011;4:31.

[28] LeRoy EC, Medsger TAJr. Criteria for the classification of early systemic sclerosis. J Rheumatol 2001;28:1573–6.

[29] van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. Radiol Phys Technol 2017;10:23–32.

[30] Depeursinge A, Chin AS, Leung AN, et al. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography. Invest Radiol 2015;50:261–7.

[31] Moon JW, Bae JP, Lee HY, et al. Perfusion- and pattern-based quantitative CT indexes using contrast-enhanced dual-energy computed tomography in diffuse interstitial lung disease: relationships with physiologic impairment and prediction of prognosis. Eur Radiol 2016;26:1368–77.

[32] Ariani A, Silva M, Bravi E, et al. Operator-independent quantitative chest computed tomography versus standard assessment of interstitial lung disease related to systemic sclerosis: a multi-centric study. Mod Rheumatol 2015;25:724–30.

[33] Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 2016;35:1207–16.

[34] Walsh SLF, Calandriello L, Silva M, et al. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. Lancet Respir Med 2018;6:837–45.

[35] Best AC, Lynch AM, Bozic CM, et al. Quantitative CT indexes in idiopathic pulmonary fibrosis: relationship with physiologic impairment. Radiology 2003;228:407–14.

[36] Solomon J, Mileto A, Nelson RC, et al. Quantitative features of liver lesions, lung nodules, and renal stones at multi-detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology 2016;279:185–94.

[37] Kim H, Park CM, Park SJ, et al. Temporal changes of texture features extracted from pulmonary nodules on dynamic contrast-enhanced chest computed tomography: how influential is the scan delay. Invest Radiol 2016;51:569–74.

[38] Larue R, van Timmeren JE, de Jong EEC, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. Acta Oncol 2017;1–0.

[39] Berenguer R, Pastor-Juan MDR, Canales-Vazquez J, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. Radiology 2018;288:407–15.

[40] Ahn SJ, Kim JH, Lee SM, et al. CT reconstruction algorithms affect histogram and texture analysis: evidence for liver parenchyma, focal solid liver lesions, and renal cysts. Eur Radiol 2018;Nov 19. doi: 10.1007/s00330-018-5829-9. [Epub ahead of print].

[41] Orlhac F, Frouin F, Nioche C, et al. Validation of a method to compensate multicenter effects affecting CT radiomics. Radiology 2019;291:53–9.

[42] Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.

[43] Prezzi D, Owczarczyk K, Bassett P, et al. Adaptive statistical iterative reconstruction (ASIR) affects CT radiomics quantification in primary colorectal cancer. Eur Radiol 2019;https://doi.org/10.1007/s00330-019-06073-3.

[44] Mannil M, von Spiczak J, Muehlematter UJ, et al. Texture analysis of myocardial infarction in CT: comparison with visual analysis and impact of iterative reconstruction. Eur J Radiol 2019;113:245–50.

[45] Peikert T, Duan F, Rajagopalan S, et al. Novel high-resolution computed tomography-based radiomic classifier for screen-identified pulmonary nodules in the National Lung Screening Trial. PLoS One 2018;13:e0196910.

[46] Hainc N, Stippich C, Stieltjes B, et al. Experimental texture analysis in glioblastoma: a methodological study. Invest Radiol 2017;52:367–73.

[47] Madan R, Matalon S, Vivero M. Spectrum of smoking-related lung diseases: imaging review and update. J Thorac Imaging 2016;31:78–91.

[48] Choi YW, Rossi SE, Palmer SM, et al. Bronchiolitis obliterans syndrome in lung transplant recipients: correlation of computed tomography findings with bronchiolitis obliterans syndrome stage. J Thorac Imaging 2003;18:72–9.

[49] Christe A, Charimo-Torrente J, Roychoudhury K, et al. Accuracy of low-dose computed tomography (CT) for detecting and characterizing the most common CT-patterns of pulmonary disease. Eur J Radiol 2013;82:e142–50.

[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.