Radiology

ORIGINAL RESEARCH • THORACIC IMAGING

Elastic Registration–driven Deep Learning for Longitudinal Assessment of Systemic Sclerosis Interstitial Lung Disease at CT

Guillaume Chassagnon, MD • Maria Vakalopoulou, PhD • Alexis Régent, MD • Mihir Sahasrabudhe, MD • Rafael Marini, MSc • Trieu-Nghi Hoang-Thi, MD • Anh-Tuan Dinh-Xuan, MD • Bertrand Dunogué, MD • Luc Mouthon, MD • Nikos Paragios, PhD • Marie-Pierre Revel, MD

From the Department of Radiology (G.C., T.N.H.T., M.P.R.), Department of Internal Medicine, Reference Center for Rare Systemic Autoimmune Diseases of Île de France (A.R., B.D., L.M.), and Department of Physiology (A.T.D.X.), Hôpital Cochin, AP-HP Centre, Université de Paris, 27 Rue du Faubourg Saint-Jacques, 75014 Paris, France; Center for Visual Computing, École CentraleSupélec, Gif-sur-Yvette, France (G.C., M.V., M.S., N.P.); and TheraPanacea, Paris, France (R.M., N.P.). Received February 5, 2020; revision requested May 4; final revision received August 17; accepted August 31. **Address correspondence to** M.P.R. (e-mail: *marie-pierre.revel@aphp.fr*).

This study was part of a research cooperation between École CentraleSupelec and Assistance Publique des Hopitaux de Paris, Cochin Hospital, which received partial financial support from GE Healthcare.

Conflicts of interest are listed at the end of this article.

See also the editorial by Verschakelen in this issue.

Radiology 2021; 298:189–198 • https://doi.org/10.1148/radiol.2020200319 • Content codes: CH CT

Background: Longitudinal follow-up of interstitial lung diseases (ILDs) at CT mainly relies on the evaluation of the extent of ILD, without accounting for lung shrinkage.

Purpose: To develop a deep learning–based method to depict worsening of ILD based on lung shrinkage detection from elastic registration of chest CT scans in patients with systemic sclerosis (SSc).

Materials and Methods: Patients with SSc evaluated between January 2009 and October 2017 who had undergone at least two unenhanced supine CT scans of the chest and pulmonary function tests (PFTs) performed within 3 months were retrospectively included. Morphologic changes on CT scans were visually assessed by two observers and categorized as showing improvement, stability, or worsening of ILD. Elastic registration between baseline and follow-up CT images was performed to obtain deformation maps of the whole lung. Jacobian determinants calculated from the deformation maps were given as input to a deep learning–based classifier to depict morphologic and functional worsening. For this purpose, the set was randomly split into training, validation, and test sets. Correlations between mean Jacobian values and changes in PFT measurements were evaluated with the Spearman correlation.

Results: A total of 212 patients (median age, 53 years; interquartile range, 45–62 years; 177 women) were included as follows: 138 for the training set (65%), 34 for the validation set (16%), and 40 for the test set (21%). Jacobian maps demonstrated lung parenchyma shrinkage of the posterior lung bases in patients found to have worsened ILD at visual assessment. The classifier detected morphologic and functional worsening with an accuracy of 80% (32 of 40 patients; 95% confidence interval [CI]: 64%, 91%) and 83% (33 of 40 patients; 95% CI: 67%, 93%), respectively. Jacobian values correlated with changes in forced vital capacity ($R = -0.38$; 95% CI: -0.25, -0.49; $P < .001$) and diffusing capacity for carbon monoxide ($R = -0.42$; 95% CI: -0.27, -0.54; $P < .001$).

Conclusion: Elastic registration of CT scans combined with a deep learning classifier aided in the diagnosis of morphologic and functional worsening of interstitial lung disease in patients with systemic sclerosis.

Online supplemental material is available for this article.

Interstitial lung disease (ILD) is a major cause of morbidity and the leading cause of death in patients with systemic sclerosis (SSc) (1,2). High-resolution CT of the chest is the most reliable method for ILD detection (3). In the large European Scleroderma Trials and Research group, the prevalence of ILD among patients with SSc at CT was 53%, whereas dyspnea was present in only 35% (4). Initial ILD extent is predictive of both disease progression and mortality (5,6). Thus, CT plays an important role in ILD screening and staging, along with pulmonary function tests (PFTs).

For disease monitoring, PFT measurements, particularly forced vital capacity predicted for age and sex, are the predominant marker. According to the American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Association criteria, in the context of idiopathic pulmonary fibrosis, functional deterioration is defined as a relative decline of 10% in absolute forced vital capacity or of 15% in absolute diffusing capacity for carbon monoxide (DLco) (7–9). However, some authors consider changes in PFTs an insensitive end point (10) and suggest combining forced vital capacity

Abbreviations

CI = confidence interval, DLco = diffusing capacity for carbon monoxide, DLco%c = percentage of predicted DLco corrected for measured hemoglobin, FVC% = percentage of predicted forced vital capacity, ILD = interstitial lung disease, IQR = interquartile range, Kco%c = percentage of predicted carbon monoxide transfer coefficient corrected for measured hemoglobin, PFT = pulmonary function test, SSc = systemic sclerosis, TLC% = percentage of predicted total lung capacity

Summary

In patients with systemic sclerosis, a deep learning classifier applied to elastic registration of chest CT images depicted lung shrinkage and functional worsening with high accuracy.

Key Results

- Worsened interstitial lung disease in patients with systemic sclerosis is associated with progressive lung parenchymal shrinkage at chest CT.
- The deep learning classifier depicted morphologic and functional worsening with an accuracy of 80% (32 of 40 patients) and 83% (33 of 40 patients), respectively.
- Jacobian values correlated with changes in forced vital capacity ($R = -0.38$; $P < .001$) and diffusing capacity for carbon monoxide ($R = -0.42$; $P < .001$).

measurement with a morphologic evaluation of disease progression as a better alternative in clinical trials (11). In patients with SSc, no clear recommendations exist regarding the role of CT in monitoring ILD. CT is often performed in situations of increased pulmonary symptoms or a decline in PFT measurements (12). Nonspecific interstitial pneumonia is the main form of SSc-related ILD, which progresses from an initial inflammatory stage to fibrosis. Thus, ILD follow-up on CT scans requires not only comparing the disease extent, but also evaluating the lung shrinkage related to the fibrosing process.

Only a few studies have evaluated longitudinal changes of SSc-related ILD at CT, and all of these have focused on the evaluation of ILD extent (6,13,14). Assessment of lung shrinkage is more complex and requires a side-by-side comparison of CT images at each level. Elastic registration of CT images provides a quantitative means to assess lung shrinkage associated with worsening during the monitoring of ILD progression. The purpose of this study was to develop a deep learning–based method to diagnose SSc-ILD worsening on the basis of lung shrinkage detection from elastic registration of chest CT scans obtained during follow-up.

Materials and Methods

Two authors (R.M. and N.P.) are employees of TheraPanacea (Paris, France). They provided support for the elastic registration process but had no control of the data or statistics.

Study Participants

This single-center retrospective study was approved by the institutional review board of the Société de Pneumologie de Langue Française (reference no. CEPRO-2017-023), which waived the need for patients' consent. Patients who met the American College of Rheumatology and the European League against Rheumatism 2013 criteria for SSc (15) were recruited from the database of the Centre de Référence des Maladies Auto-Immunes Systémiques Rares d'Ile de France at Cochin Hospital. This database collects the sociodemographic, clinical, morphologic, and biologic characteristics as well as PFT measurements of patients with SSc referred to this tertiary care center. All consecutive patients with SSc evaluated at our university hospital between January 2009 and October 2017 were eligible. The inclusion criterion was the availability of at least two consecutive unenhanced CT scans of the chest obtained in the supine position. Exclusion criteria were *(a)* presence of motion artifacts, *(b)* signs of another lung disease or acute complication at CT, and *(c)* unavailability of PFTs within 3 months of the CT scans (Fig 1). If patients had more than two CT scans meeting the inclusion criteria, only the oldest and the more recent CT scans were considered.

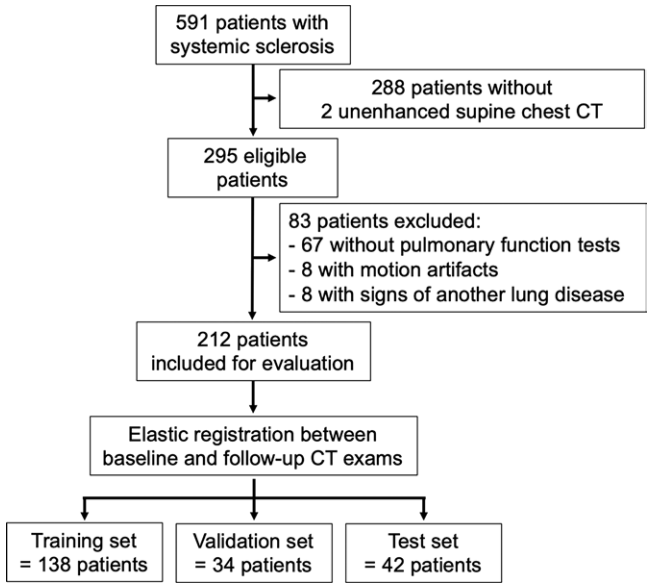

Figure 1: Flowchart of study patients.

CT Scans

Inspiratory unenhanced CT scans of the whole lung were acquired with multisection CT devices (Somatom Sensation 16, Somatom DS, and Somatom AS+; Siemens Healthineers, Erlangen, Germany; and Revolution HD; GE Healthcare, Milwaukee, Wis). Acquisition and reconstruction parameters were those fulfilling CT criteria. Images were reconstructed in the axial plane with a section thickness of 0.625–1.25 mm using filtered back projection (384 of 424 CT examinations [91%]) or iterative reconstruction algorithms (40 of 424 CT examinations [9%]) (sinogram-affirmed iterative reconstruction level 3 or adaptive statistical iterative reconstruction-V, 70%), and a high-frequency kernel (B70f, I70f, or Lung).

Visual Analysis

Visual image analysis was performed by two chest radiologists in consensus (G.C. and M.P.R., with 4 years and 18 years of experience in chest imaging, respectively). The two radiologists were blinded to the PFT measurements and the results of the deep learning classifier. CT scans were reviewed for the pres-

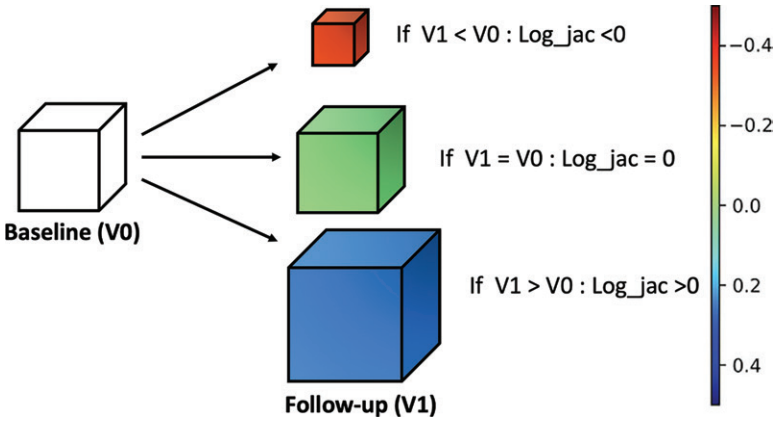

Figure 2: Illustration of Jacobian determinant (quantitative value of deformation matrix of each voxel). If voxel size remains similar after deformation (V0 = V1), logarithm of Jacobian determinant (Log_Jac) is 0. If deformed voxel is smaller than original, Log_Jac is negative, whereas if deformed voxel is larger, Log_Jac is positive.

ence of ILD, defined by the presence of ground-glass opacities, reticulations, traction bronchiectasis, or honeycombing in various combinations. Then, a side-by-side comparison of baseline and follow-up CT images was performed in consensus by the same two radiologists to assess whether ILD showed morphologic stability, worsening, or improvement. CT scans were considered to show improvement when they were associated with a decrease in volume of ground-glass opacities. Worsening consisted of an increase in extent of ILD or an increase in traction bronchiectasis and/or honeycombing with or without an increase in the extent of ILD.

PFT Measurements

PFTs were measured within 3 months of the baseline and follow-up chest CT and were retrieved from the database. PFT measurements included percentage of predicted forced vital capacity (FVC%), percentage of predicted total lung capacity (TLC%), percentage of predicted DLco corrected for measured hemoglobin (DLco%c), and percentage of predicted carbon monoxide transfer coefficient corrected for measured hemoglobin (Kco%c). Functional changes were evaluated according to the American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Association criteria. Functional worsening was defined by a decrease in absolute FVC% of 10% or a decrease in absolute DLco% of 15% (7).

Elastic Registration

Image preprocessing was performed with Python software (version 2.7; Python Software Foundation, Wilmington, Del). The preprocessing steps included image resampling to a 1-mm isotropic resolution and lung segmentation using an in-house deep learning–based segmentation tool. Follow-up CT images were elastically registered to match baseline images and to calculate the deformation maps. Elastic registration was performed using a robust, multimetric, multimodal graph-based registration algorithm (16,17). Jacobian maps were obtained by calculating the logarithm of the Jacobian (or log_jac) determinant for each voxel of the deformation matrix. The Jacobian determinant is a quantitative measurement of the deformation applied to each voxel to have the baseline lung scan matching the follow-up examination. It quantifies the stretching or shrinkage of each voxel. If the voxel size remains similar after deformation, log_jac is 0. If the deformed voxel is smaller than the original, log_jac is negative; if the deformed voxel is larger, log_jac is positive (Fig 2).

To obtain statistics for all patients and allow comparison between groups, all Jacobian maps were elastically registered to a common template using the same registration algorithm (16,17).

Jacobian maps were visually compared between patients with functional stability or functional worsening and between patients with improvement, stability, or worsening based on visual morphologic assessment. The registered Jacobian maps were averaged for each patient group to produce a unique three-dimensional Jacobian map for visual comparison of each group. Mean log_jac values of each patient were also calculated for comparison between groups and evaluation of the correlation with PFT changes.

Deep Learning–based Classification

For automated detection of disease worsening at CT, a deep learning architecture (fully connected convolutional neural network with four layers and cross entropy loss) was trained twice, successively targeting depiction of functional or morphologic changes (Figure E1 [online]). The first model was trained to depict functional worsening—as determined from the PTF scores—whereas the other model was trained to depict morphologic worsening—as determined from the radiologists' (G.C. and M.P.R.) assessment. Patients showing visual improvement were included in the same group as those with stable disease. The Jacobian maps of the whole lung were used as input, and patients were randomly split into a learning set (172 patients; 80% for training and 20% for validation) and a test set (40 patients). Patient characteristics in the learning set (172 patients) and test set (40 patients) were not statistically different ($P \geq .17$; Table E1 [online]). More information regarding the deep learning model can be found in Appendix E1 (online). The code of the deep learning classifiers is available at *https://github.com/msahasrabudhe/jacs_clf*.

Quantification of ILD Extent

For comparison purposes, the extent of ILD was quantified on each CT scan using a previously reported deep learning–based method developed for SSc-ILD segmentation (18,19). Through use of this method, ILD extent was expressed as the percentage of diseased lung from the total lung volume.

Statistical Analysis

Statistical analysis was performed with R software (version 3.3.3; the R Foundation for Statistical Computing, Vienna, Austria). Characteristics of the study participants were compared using the Fisher exact test for categorical variables or analysis of variance (with the Scheffe method for post hoc anal-

Table 1: Patient Characteristics according to Functional Changes

Characteristic	All Patients (n = 212)	Patients with Functional Worsening (n = 91)	Patients with Functional Stability (n = 121)	P Value
Median age (y)*	53 (45–62)	54 (48–63)	52 (42–62)	.12
Women	177 (83)	79 (87)	98 (81)	.35
Limited cutaneous systemic sclerosis	128 (60)	51 (56)	77 (64)	.32
Median Rodnan score at baseline*	6 (2–13)	7 (2–14)	6 (2–13)	.65
Median baseline pulmonary function test values*				
TLC%	95 (81–108)	94 (78–109)	96 (87–107)	.33
FVC%	91 (75–104)	91 (74–106)	93 (75–103)	.61
DLco%c	63 (49–76)	62 (46–75)	65 (50–77)	.37
Kco%c	79 (68–91)	79 (67–90)	80 (70–91)	.78
Median interval between baseline and follow-up chest CT (mo)*	37 (23–53)	45 (34–65)	27 (16–49)	<.001
ILD at baseline CT	147 (69)	71 (78)	76 (63)	.02
ILD extent at baseline CT (%)	6.1 (0.4, 19.1)	9.3 (1.0, 25.0)	3.7 (0.3, 17.2)	.048
Median changes in pulmonary function test results*				
TLC%	0 (−7 to 5)	−4 (−10 to 3)	1 (−3 to 6)	.01
FVC%	−2 (−8 to 6)	−8 (−12 to −2)	3 (−3 to 7)	<.001
DLco%c	−4 (−11 to 0)	−13 (−18 to −8)	−2 (−5 to 3)	<.001
Kco%c	−6 (−14 to 2)	−13 (−22 to −5)	−2 (−7 to 5)	<.001
Morphologic worsening at CT	73 (34)	47 (52)	26 (21)	<.001
Median changes in ILD extent	0.1 (−0.8, 3.6)	1.5 (0.0, 7.8)	0 (−2.2, 0.8)	<.001
Median mean log_jac*	0.02 (−0.05 to 0.07)	0.05 (0.00–0.11)	0.01 (−0.08 to 0.03)	<.001

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. DLco%c = percentage of predicted diffusing capacity for carbon monoxide, FVC% = percentage of predicted forced vital capacity, ILD = interstitial lung disease, Kco%c = percentage of predicted carbon monoxide transfer coefficient corrected for measured hemoglobin, log_jac = logarithm of the Jacobian determinant, TLC% = percentage of predicted total lung capacity.

* Numbers in parentheses are interquartile ranges.

ysis) and Mann-Whitney test for quantitative data. Correlation between mean log_jac values and changes in PFT parameters (TLC%, FVC%, DLco%c, and Kco%c) was evaluated using the Spearman correlation coefficient. Receiver operating characteristic curve analyses were used to find the best thresholds for identifying patients with functional and morphologic worsening according to ILD extent changes in the training and/or validation data set. Sensitivity, specificity, and accuracy of the classifiers to depict morphologic or functional worsening were calculated using visual analysis of CT scans or functional changes as a reference. $P < .05$ was considered to indicate a statistically significant difference.

Results

Patient Characteristics

During the study period, 591 patients with SSc from our reference center underwent chest CT at our radiology department. Among the 295 patients who had undergone at least two unenhanced CT examinations of the chest performed in the supine position, 83 were excluded because of unavailability of PFTs (n = 67), motion artifacts (n = 8), or signs of another lung disease (n = 8); 212 were included (Fig 1). Most patients were women (n = 177 [83%]) and had limited cutaneous disease (n = 128 [60%]) (Table 1).

At baseline, the median age was 53 years (interquartile range [IQR], 45–62 years), and 69% (147 of 212) had ILD at CT. The median extent of ILD at chest CT was 6.1% of the lung volume (IQR, 0.4%–19.1%). The median FVC% was 91% (IQR, 75%–104%); 18% of patients (63 of 212) had a decreased FVC% (<80% of the predicted value according to Goh et al [5]). The DLco measurement at baseline was available for 196 patients (92%). The median DLco%c was 63% (IQR, 49%–76%); 61% of patients (120 of 196) had a decreased DLcoc% (<70% of the predicted value).

The median interval between baseline and follow-up was 37 months (IQR, 23–53 months). It was longer for patients with functional worsening (median, 45 months [IQR, 34–65 months]) than for patients with functional stability (median, 27 months [IQR, 16–46 months]) ($P < .001$) (Table 1). It was also longer in patients showing morphologic worsening at CT (median, 43 months [IQR, 32–59 months]) compared with those having morphologic stability (median, 32 months [IQR, 19–49 months]) ($P = .003$) at visual assessment (Table 2).

Regarding PFTs, all but 25 patients had both baseline and follow-up measurement of DLcoc%. Overall, we observed a mild decline in pulmonary function, with a median change of −2% (IQR, −8% to 6%) in FVC% and −4% (IQR, −11% to 0%) in DLco%c (Table 1). Ninety-one of the 212 patients (43%) met the American Thoracic Society, European Respiratory

Table 2: Patient Characteristics according to Morphologic Changes

Characteristic	All Patients (n = 212)	Patients with Visual Worsening at CT (n = 73)	Patients with Visual Stability at CT (n = 132)	Patients with Visual Improvement at CT (n = 7)	P Value
Median age (y)*	53 (45–62)	51 (45–63)	54 (44–61)	57 (55–63)	.38
Women	177 (83)	58 (79)	114 (86)	5 (71)	.21
Limited cutaneous systemic sclerosis	128 (60)	33 (45)	92 (70)	3 (43)	.001
Median Rodnan score at baseline*	6 (2–13)	10 (4–10)	4 (5–19)	3 (0–20)	<.001
Median baseline pulmonary function test values*					
TLC%	95 (81–108)	90 (75–101)	99 (87–110)	83 (78–97)	<.001
FVC%	91 (75–104)	83 (66–100)	94 (81–105)	82 (75–88)	.005
DLco%c	63 (49–76)	56 (42–70)	69 (53–80)	46 (36–56)	<.001
Kco%c	79 (68–91)	76 (64–87)	82 (71–92)	75 (66–84)	.14
Median interval between baseline and follow-up chest CT (mo)*	37 (23–53)	43 (32–59)	32 (19–49)	46 (24–65)	.003
ILD at baseline CT	147 (69)	66 (90)	74 (56)	7 (100)	<.001
ILD extent at baseline CT (%)	6.1 (0.4, 19.1)	13.2 (3.5, 25.3)	2.3 (0.2, 15.0)	17.7 (8.6, 24.1)	.01
Median changes in pulmonary function test results*					
TLC%	0 (−7 to 5)	−3 (−10 to 3)	0 (−5 to 6)	6 (−2 to 13)	.01
FVC%	−2 (−8 to 6)	−7 (−11 to 0)	2 (−5 to 6)	15 (5–22)	<.001
DLco%c	−4 (−11 to 0)	−10 (−16 to −4)	−3 (−9 to 1)	−4 (−8 to 11)	<.001
Kco%c	−6 (−14 to 2)	−10 (−20 to −1)	−3 (−11 to 4)	−6 (−9 to 1)	.001
Functional deterioration according to ATS/ERS/JRS/ALAT guidelines	91 (43)	47 (64)	41 (31)	3 (43)	<.001
Median changes in ILD extent	0.1 (−0.8, 3.6)	4.5 (0.8, 8.9)	0 (−1.2, 0.5)	−6.6 (−4.7, −17.2)	<.001
Median mean log_jac*	0.02 (−0.05 to 0.07)	0.05 (−0.01 to 0.10)	0.01 (−0.07 to 0.04)	−0.08 (−0.15 to −0.00)	.005

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. ATS/ERS/JRS/ALAT = American Thoracic Society/European Respiratory Society/Japanese Respiratory Society/Latin American Thoracic Association, DLco%c = percentage of predicted diffusing capacity for carbon monoxide, FVC% = percentage of predicted forced vital capacity, ILD = interstitial lung disease, Kco%c = percentage of predicted carbon monoxide transfer coefficient corrected for measured hemoglobin, log_jac = logarithm of the Jacobian determinant, TLC% = percentage of predicted total lung capacity.

* Numbers in parentheses are interquartile ranges.

Society, Japanese Respiratory Society, and Latin American Thoracic Association criteria for functional deterioration. Visual comparison between baseline and follow-up CT scans showed morphologic worsening in 73 of the 212 patients (34%), including eight patients who developed an ILD. In seven of the 212 patients (3%), ILD had improved on the follow-up CT scan, whereas in the remaining 132 patients (62%), the disease was considered morphologically stable (Table 2). The seven patients who improved received steroids and immunosuppressive treatments in the time between the two CT scans.

Morphologic worsening was more frequent in patients with functional worsening (47 of 91 [52%]) than in patients with functional stability (26 of 121 [21%]) ($P < .001$) (Table 1). Similarly, PFT changes differed between patients showing morphologic worsening, stability, or improvement ($P \leq .001$) (Table 2). The post hoc tests showed that decline in TLC%, FVC%, DLco%c, and Kco%c was more frequent for patients with visual worsening at CT than for patients with visual stability (TLC%: −3% [IQR: −10% to 3%] vs 0% [IQR: −5% to 6%], respectively, $P = .03$; FVC%: −7% [IQR: −11% to 0%] vs 2% [IQR: −5% to 6%], $P = .001$; DLco%c: −10% [IQR: −16% to −4%] vs −3% [IQR: −9% to 1%], $P < .001$; and Kco%c: −10% [IQR: −20% to −1%] vs −3% [IQR: −11% to 4%], $P = .001$) (Figs 3, 4). Similarly, the post hoc tests showed that the decline in FVC% and DLco%c was more frequent for patients with morphologic worsening compared with those with improvement at CT (FVC%: −7% [IQR: −11% to 0%] vs 15% [IQR: 5%–22%], $P = <.001$; DLco%c: −10% [IQR: −16% to −4%] vs −4% [IQR: −8% to 11%], $P = .02$), whereas median FVC% increased for patients with morphologic improvement in comparison with patients having morphologic stability at CT (median change, 15% [IQR: 5%–22%] vs 2% [IQR: −5% to 6%], respectively, $P = .001$) (Table 2). Both median ILD extent at baseline and median absolute changes in ILD extent were significantly higher for patients with functional worsening than for those with stable disease (ILD extent: 9.3% [IQR: 1.0%–25.0%] vs 3.7% [IQR: 0.3%–17.2%], respectively, $P = .048$; median absolute change in ILD extent: 1.5% [IQR: 0.0%–7.8%] vs 0% [IQR: −2.2% to 0.8%], $P < .001$) (Table 1). Median ILD extent and absolute changes in ILD extent also significantly differed between patients showing morphologic worsening, stability, or improvement ($P = .01$ and $P \leq .001$,

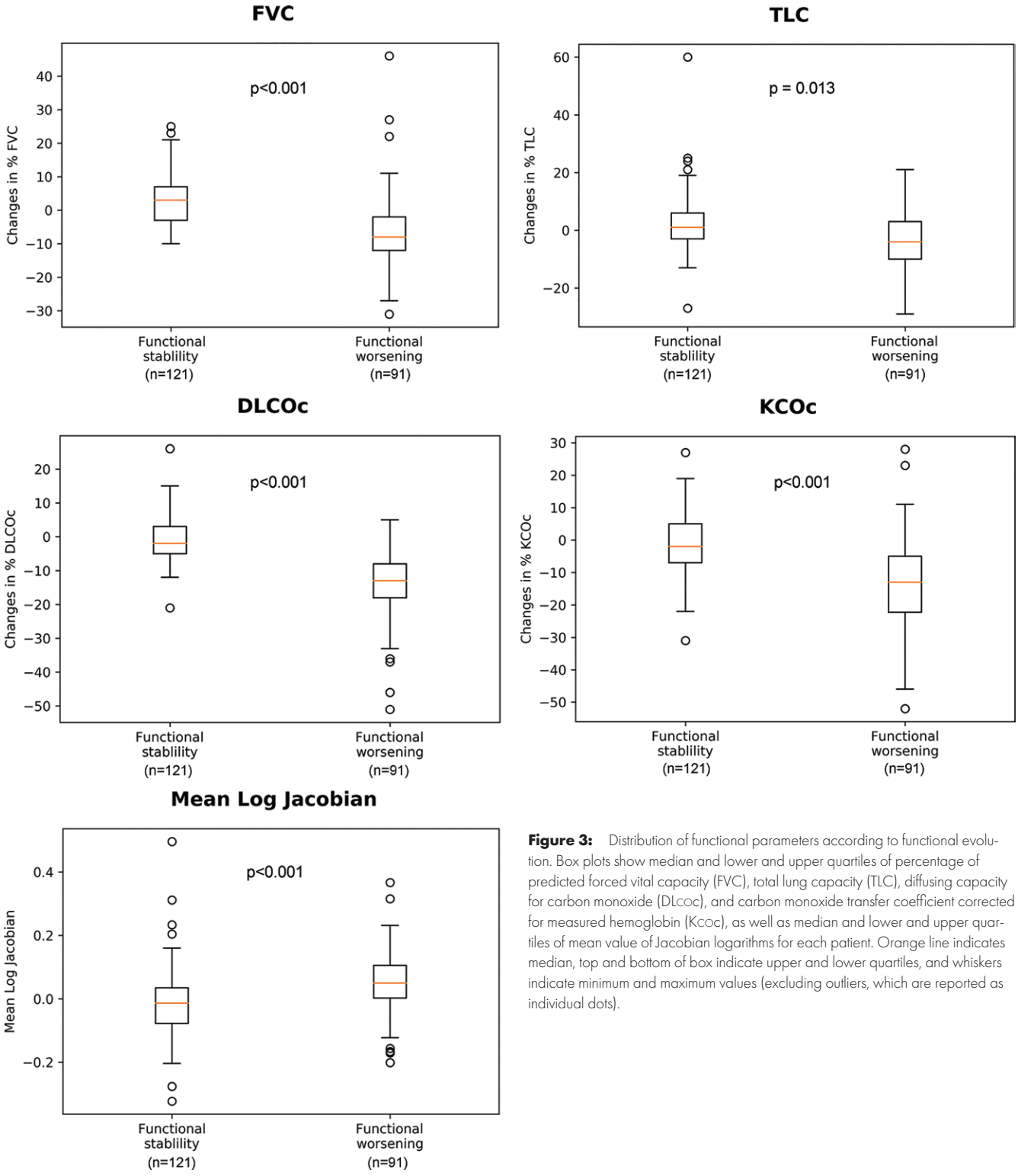

Figure 3: Distribution of functional parameters according to functional evolution. Box plots show median and lower and upper quartiles of percentage of predicted forced vital capacity (FVC), total lung capacity (TLC), diffusing capacity for carbon monoxide (DLCOc), and carbon monoxide transfer coefficient corrected for measured hemoglobin (KCOc), as well as median and lower and upper quartiles of mean value of Jacobian logarithms for each patient. Orange line indicates median, top and bottom of box indicate upper and lower quartiles, and whiskers indicate minimum and maximum values (excluding outliers, which are reported as individual dots).

respectively) (Table 2). The post hoc tests showed a higher ILD extent at baseline and a higher increase in ILD extent for patients with morphologic worsening compared with those with stable disease (ILD extent at baseline: 13.2% [IQR: 3.5%–25.3%] vs 2.3% [IQR: 0.2%–15.0%], respectively, P = .02; increase in ILD extent: 4.5% [IQR: 0.8%–8.9%] vs 0 [IQR: −1.2% to 0.5%], $P < .001$) (Table 2).

Elastic Registration

Mean log_jac of the lung significantly differed between patient groups according to functional changes (worsening group: 0.05 [IQR: 0.00–0.11]; stability group: 0.01 [IQR: −0.08 to 0.03], $P < .001$) (Fig 3) or morphologic changes (worsening group: 0.05 [IQR: −0.01 to 0.10]; stability group: 0.01 [IQR: −0.07 to 0.04]; improvement group: −0.08 [IQR: −0.15 to

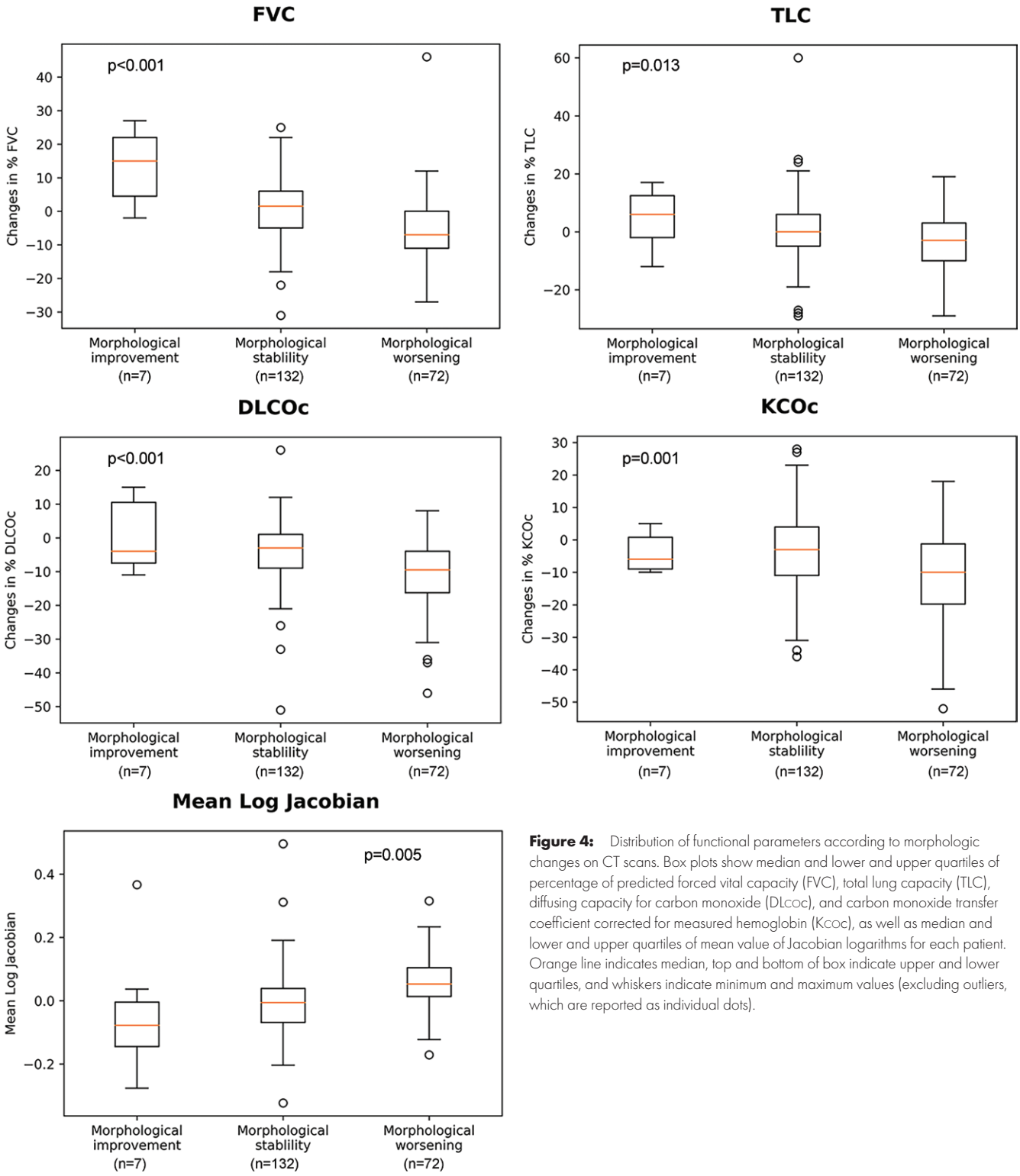

Figure 4: Distribution of functional parameters according to morphologic changes on CT scans. Box plots show median and lower and upper quartiles of percentage of predicted forced vital capacity (FVC), total lung capacity (TLC), diffusing capacity for carbon monoxide (DLco), and carbon monoxide transfer coefficient corrected for measured hemoglobin (Kco), as well as median and lower and upper quartiles of mean value of Jacobian logarithms for each patient. Orange line indicates median, top and bottom of box indicate upper and lower quartiles, and whiskers indicate minimum and maximum values (excluding outliers, which are reported as individual dots).

−0.00], $P = .005$) (Fig 4). Patients experiencing disease worsening (morphologic or functional) had higher values of mean log_jac of the lung.

We found a moderate negative correlation between individual mean log_jac value and changes in DLco%c ($R = -0.42$; 95% CI: −0.27, −0.54; $P < .001$) and a weak negative correlation between individual mean log_jac value and changes in FVC% ($R = -0.38$; 95% CI: −0.25, −0.49; $P < .001$), TLC% ($R = -0.27$; 95% CI: −0.13, −0.39; $P < .001$), and Kco%c ($R = -0.30$; 95% CI: −0.15, −0.43; $P < .001$).

The qualitative assessment of the distribution of the lung log_jac values produced clear visual separation between the patient groups. Sagittal views summarizing the lung distribution of log_jac values in each group showed homogeneous log_jac

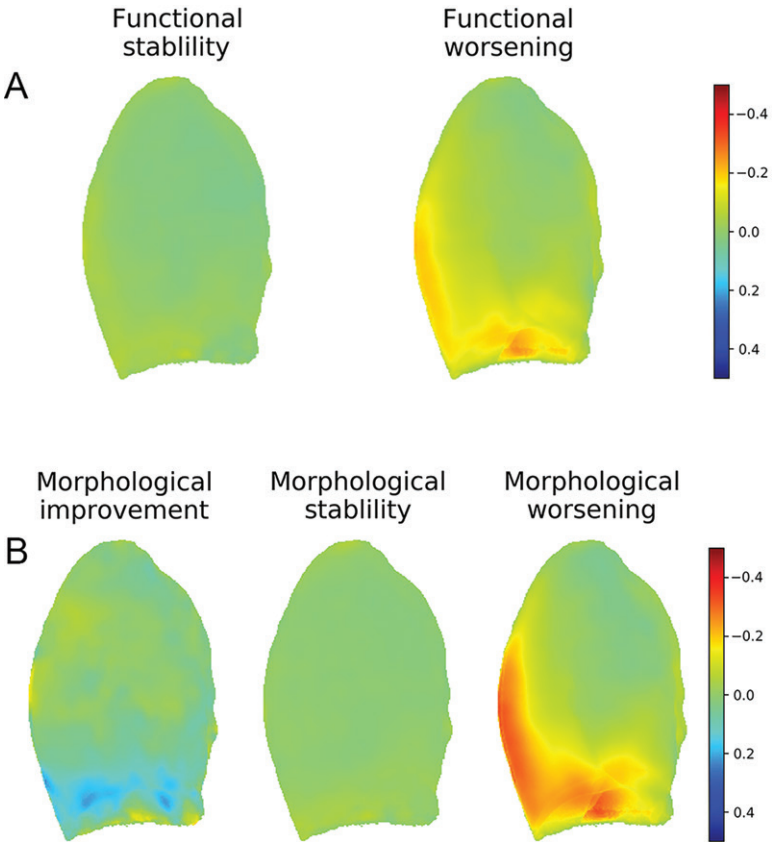

Figure 5: Sagittal projection of mean of Jacobian determinant maps. *A*, Two groups of patients according to functional changes. *B*, Three groups of patients according to morphologic changes.

values in patients with stable disease on the basis of functional or morphologic criteria (Fig 5). Conversely, lung shrinkage (negative log_jac values) was seen in the posterior part of the lung bases in patients having functional or morphologic worsening, whereas the same areas showed expansion (positive log_jac values) in the small group of patients with morphologic improvement on native CT images. Representative CT images of progressive and stable ILD along with Jacobian maps are shown in Figure 6.

Depiction of Disease Worsening

With use of American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Association functional deterioration criteria as a reference standard to train and test the classifier, lung shrinkage assessment reached a sensitivity of 89% (16 of 18; 95% CI: 65%, 99%), a specificity of 77% (17 of 22; 95% CI: 55%, 92%), and an accuracy of 83% (33 of 40; 95% CI: 67%, 93%) to depict disease worsening (16 true-positive results, 17 true-negative results, five false-positive results, and two false-negative results). Positive and negative predictive values were 76% (16 of 21; 95% CI: 53%, 92%) and 89% (17 of 19; 95% CI: 67%, 99%), respectively. With use of visual assessment of morphologic changes as a reference standard to train and test the model, lung shrinkage assessment obtained a sensitivity of 79% (11 of 14; 95% CI: 49%, 95%), a specificity of 81% (21 of 26; 95% CI: 60%, 93%), and an accuracy of 80% (32 of 40; 95% CI: 64%, 91%) (11 true-positive results, 21 true-negative results, five false-positive results, and three false-negative results). Positive and negative predictive values were 69% (11 of 16; 95% CI: 41%, 89%) and 88% (21 of 24; 95% CI: 68%, 97%), respectively.

With use of increase of ILD extent, quantified using deep learning, as a worsening criterion, the best thresholds for identifying functional and morphologic worsening were 0.82% and 0.45%, respectively, in the training and validation sets. On the basis of these thresholds, the increase in ILD extent had a sensitivity of 61% (11 of 18; 95% CI: 36%, 83%), a specificity of 73% (16 of 22; 95% CI: 50%, 89%), and an accuracy of 68% (27 of 40; 95% CI: 51%, 81%) to depict functional worsening in the test set. For morphologic worsening, sensitivity, specificity, and accuracy were 71% (10 of 14; 95% CI: 42%, 92%), 65% (17 of 26; 95% CI: 44%, 83%), and 68% (27 of 40; 95% CI: 51%, 81%), respectively.

Discussion

Lung shrinkage is an important component of worsening lung fibrosis but is difficult to assess on the basis of visual assessment. We found that elastic registration applied to CT images in combination with a deep learning classifier depicted lung shrinkage and helped assess morphologic and functional worsening with an accuracy of 80% and 83%, respectively. Jacobian maps demonstrated lung parenchyma shrinkage of the posterior lung bases in patients found to have worsened interstitial lung disease at visual assessment, but not in patients with stable or improving disease. The 80%–83% accuracy obtained for depicted disease worsening was based only on an evaluation of the lung deformation, without taking into account an increase in disease extent, which is also a factor of worsening (20,21) and may not be associated with lung shrinkage in the early phase of fibrosis. The approach based on disease extent evaluation as a worsening criterion had lower performance in our study.

In PFTs, lung fibrosis not only decreases lung volumes but is also associated with a decrease in DLco and carbon monoxide transfer coefficient. These parameters correlate to ILD extent at CT in both SSc and idiopathic pulmonary fibrosis (5,21–26). Both forced vital capacity and DLco have been long been used to monitor ILD progression in SSc (27). We observed a decrease of these functional parameters in patients with visual worsening at CT. The negative correlations between the mean log_jac values and changes in PFT parameters ($R = -0.38$ for FVC%, $R = -0.27$ for TLC%, and $R = -0.42$ for DLco%c), although only weak to moderate, were in the range of those previously reported for fibrotic ILDs, using other automated methods based on imaging (20,21). In patients with SSc, Kim et al (20) reported a weak correlation between changes in disease extent and both forced vital capacity ($R = -0.33$ to -0.40) and total lung capacity

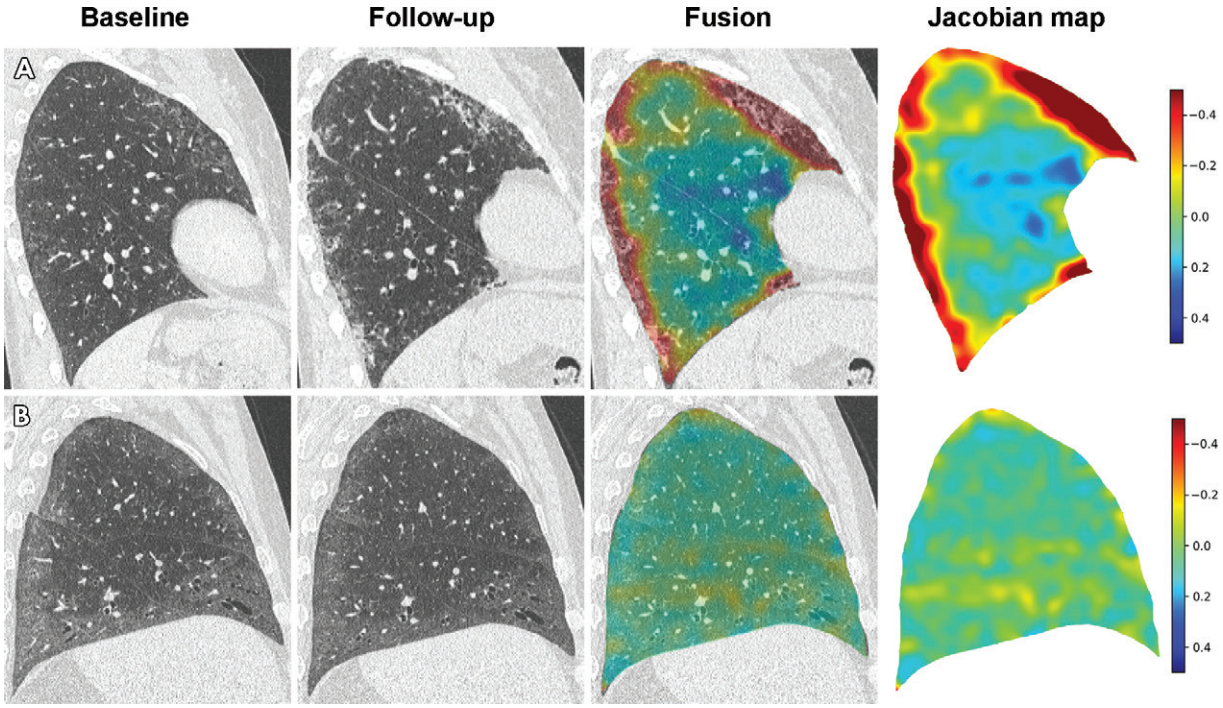

Figure 6: CT images representative of progressive and stable interstitial lung disease (ILD) along with Jacobian maps. *A*, Patient with progressive ILD. Follow-up chest CT scan at 53 months shows increase of subpleural reticulations and ground-glass opacities. Jacobian map shows increased Jacobian determinant logarithm in these areas, corresponding to focal lung shrinkage. This patient presented both functional and morphologic worsening, correctly identified with deep learning models. *B*, Patient with stable ILD. Follow-up chest CT scan at 6 months shows stable ILD with no increase of ground-glass opacities or traction bronchiectasis in lingula. Jacobian map shows Jacobian determinant logarithm close to zero, consistent with absence of focal lung shrinkage. Functional and morphologic stability were correctly identified with deep learning models.

(R = −0.16 to −0.18). In patients with idiopathic pulmonary fibrosis, Humphries et al (21) reported a weak-to-moderate inverse correlation between changes in disease extent and changes in DLco% (R = −0.27 to −0.54) and in FVC% (R = −0.43 to −0.44). The weak correlation between mean log_jac of the lung and changes in total lung capacity suggests that evaluating the deformation of the whole lung is not an optimal marker to quantify shrinkage due to lung fibrosis. PFTs help detect a restrictive pattern but are based on an overall assessment of lung volumes, which is not sensitive to focal worsening. Elastic registration could determine and characterize lung changes locally. The main contribution of Jacobian maps is to provide spatial information about the lung deformation. Averaging Jacobian determinants of the whole lung volume negatively affects the correlation to functional changes because some lung areas expand during the image registration process to compensate for shrinkage areas.

With the calculation of Jacobian determinant maps, we demonstrated that lung shrinkage predominates in the posterior areas of lung bases, the predominant location of ILD in patients with SSc. In patients with visually stable disease at CT, such shrinkage was not observed, whereas Jacobian maps showed an expansion of the same areas in the few patients with morphologic improvement. Another study that used a conceptually similar registration approach at lung MRI (28) reported that the posterior areas of lung bases showed the most important deformation between inspiration and expiration in healthy individuals, which was lost in patients with SSc and lung fibrosis.

Our study had limitations. First, some of the follow-up CT scans may have been obtained to explore pulmonary function decline due to causes other than fibrosis, which may have underestimated the correlation between functional deterioration and lung deformation. Second, because of the retrospective design, the range of the time interval between baseline and follow-up CT scans was uneven and shorter in patients with stable disease. Third, functional deterioration in SSc is not always due to pulmonary fibrosis. Patients can develop pulmonary hypertension, which cannot be depicted with an analysis of lung deformation. This was not accounted for in our study. Fourth, we evaluated CT performed in clinical routine, without respiratory gating, which could induce variability regarding the quality of deep inspiration. However, we can assume that patients with SSc having repeated follow-up evaluation, including PFTs, were "trained" to perform respiratory maneuvers, which probably minimized the variability. Finally, we did not account for worsening of skin sclerosis as a factor of lung function deterioration. High Rodnan scores (>20) can result in decreased chest expansion (29) but would not affect the lung bases, as opposed to lung shrinkage depicted on lung deformation maps.

In conclusion, lung shrinkage detected with elastic registration of CT scans combined with a deep learning classifier can be used to assess the worsening of interstitial lung disease in systemic sclerosis.

Author contributions: Guarantors of integrity of entire study, G.C., M.P.R.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual

content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, G.C., M.V., A.R., A.T.D.X., M.P.R.; clinical studies, G.C., A.R., T.N.H.T., A.T.D.X., B.D.; experimental studies, G.C., M.V., M.S., R.M., T.N.H.T.; statistical analysis, G.C., M.V., N.P.; and manuscript editing, G.C., M.V., A.R., M.S., A.T.D.X., B.D., L.M., N.P., M.P.R.

Disclosures of Conflicts of Interest: G.C. Activities related to the present article: institution received grant from GE Healthcare. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **M.V.** disclosed no relevant relationships. **A.R.** disclosed no relevant relationships. **M.S.** disclosed no relevant relationships. **R.M.** disclosed no relevant relationships. **T.N.H.T.** disclosed no relevant relationships. **A.T.D.X.** disclosed no relevant relationships. **B.D.** disclosed no relevant relationships. **L.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is a consultant for Janssen-Cilag; received reimbursement for travel, accommodations, and meeting expenses from LFB, Shire Grifols. Other relationships: disclosed no relevant relationships. **N.P.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: receives payment from Artedrone for board membership; has grant/grants pending; receives payment for lectures, including service on speakers bureaus, from AstraZeneca and Ipsen; receives royalties from École CentraleSupelec; holds stock/stock options in TheraPanacea; receives payment from École CentraleSupelec for intellectual property. **M.P.R.** disclosed no relevant relationships.

References

1. Rubio-Rivas M, Royo C, Simeón CP, Corbella X, Fonollosa V. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. Semin Arthritis Rheum 2014;44(2):208–219.
2. Tyndall AJ, Bannert B, Vonk M, et al. Causes and risk factors for death in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. Ann Rheum Dis 2010;69(10):1809–1815.
3. Suliman YA, Dobrota R, Huscher D, et al. Brief report: pulmonary function tests: high rate of false-negative results in the early detection and screening of scleroderma-related interstitial lung disease. Arthritis Rheumatol 2015;67(12):3256–3261.
4. Meier FMP, Frommer KW, Dinser R, et al. Update on the profile of the EUSTAR cohort: an analysis of the EULAR Scleroderma Trials and Research group database. Ann Rheum Dis 2012;71(8):1355–1360.
5. Goh NSL, Desai SR, Veeraraghavan S, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. Am J Respir Crit Care Med 2008;177(11):1248–1254.
6. Moore OA, Goh N, Corte T, et al. Extent of disease on high-resolution computed tomography lung is a predictor of decline and mortality in systemic sclerosis-related interstitial lung disease. Rheumatology (Oxford) 2013;52(1):155–160.
7. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183(6):788–824.
8. Wells AU. Interstitial lung disease in systemic sclerosis. Presse Med 2014;43(10 Pt 2):e329–e343.
9. Le Gouellec N, Duhamel A, Perez T, et al. Predictors of lung function test severity and outcome in systemic sclerosis-associated interstitial lung disease. PLoS One 2017;12(8):e0181692.
10. Wells AU. Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow's ear. Thorax 2013;68(4):309–310.
11. Khanna D, Seibold J, Goldin J, Tashkin DP, Furst DE, Wells A. Interstitial lung disease points to consider for clinical trials in systemic sclerosis. Rheumatol Oxf Engl. 2017;56(suppl_5):v27–v32.
12. Wallace B, Vummidi D, Khanna D. Management of connective tissue diseases associated interstitial lung disease: a review of the published literature. Curr Opin Rheumatol 2016;28(3):236–245.
13. Goldin J, Elashoff R, Kim HJ, et al. Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. Chest 2009;136(5):1333–1340.
14. Tashkin DP, Roth MD, Clements PJ, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. Lancet Respir Med 2016;4(9):708–719.
15. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum 2013;65(11):2737–2747.
16. Ferrante E, Dokania PK, Marini R, Paragios N. Deformable registration through learning of context-specific metric aggregation. In: Wang Q, Shi Y, Suk HI, Suzuki K, eds. Machine Learning in Medical Imaging. MLMI 2017. Lecture Notes in Computer Science, vol 10541. Cham, Switzerland: Springer, 2017; 256–265.
17. Glocker B, Sotiras A, Komodakis N, Paragios N. Deformable medical image registration: setting the state of the art with discrete methods. Annu Rev Biomed Eng 2011;13(1):219–244.
18. Vakalopoulou M, Chassagnon G, Bus N, et al. AtlasNet: Multi-atlas Non-linear Deep Networks for Medical Image Segmentation. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, eds. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11073. Cham, Switzerland: Springer, 2018; 658–666.
19. Chassagnon G, Vakalopolou M, Régent A, et al. Deep learning–based approach for automated assessment of interstitial lung disease in systemic sclerosis on CT images. Radiol Artif Intell 2020;2(4):e190006.
20. Kim HJ, Brown MS, Elashoff R, et al. Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. Eur Radiol 2011;21(12):2455–2465.
21. Humphries SM, Swigris JJ, Brown KK, et al. Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. Eur Respir J 2018;52(3):1801384.
22. Diot E, Boissinot E, Asquier E, et al. Relationship between abnormalities on high-resolution CT and pulmonary function in systemic sclerosis. Chest 1998;114(6):1623–1629.
23. Ooi GC, Mok MY, Tsang KWT, et al. Interstitial lung disease in systemic sclerosis. Acta Radiol 2003;44(3):258–264.
24. Camiciottoli G, Orlandi I, Bartolucci M, et al. Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. Chest 2007;131(3):672–681.
25. Kim HG, Tashkin DP, Clements PJ, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 2010;28(5 Suppl 62):S26–S35.
26. Jacob J, Bartholmai BJ, Rajagopalan S, et al. Automated quantitative computed tomography versus visual computed tomography scoring in idiopathic pulmonary fibrosis: validation against pulmonary function. J Thorac Imaging 2016;31(5):304–311.
27. Caron M, Hoa S, Hudson M, Schwartzman K, Steele R. Pulmonary function tests as outcomes for systemic sclerosis interstitial lung disease. Eur Respir Rev 2018;27(148):170102.
28. Chassagnon G, Martin C, Marini R, et al. Use of elastic registration in pulmonary MRI for the assessment of pulmonary fibrosis in patients with systemic sclerosis. Radiology 2019;291(2):487–492.
29. Rodnan GP, Lipinski E, Luksick J. Skin thickness and collagen content in progressive systemic sclerosis and localized scleroderma. Arthritis Rheum 1979;22(2):130–140.