Reumatismo - The Italian Journal of Rheumatology

Reumatismo 2026; 78:1920
doi:10.4081/reumatismo.2026.1920

Computed tomography of interstitial lung disease in systemic sclerosis: dataset and deep learning model for pulmonary lesion segmentation

Amir M. Vahdani,[1] Aida Mohamadi,[2] Sepehr Nayebirad,[3] Shayan Forghani,[1] Elaheh Karimi,[1] Yousef Vojgani,[2] Zahra Tamartash,[2] Mehrnaz Asadi,[4] Leila Aghaghazvini,[5] Farhad Gharibdoost,[2] Hoda Kavosi[2]

[1]School of Medicine, Tehran University of Medical Sciences; [2]Rheumatology Research Center, Tehran University of Medical Sciences; [3]Tehran Heart Center, Cardiovascular Diseases Research Institute, Tehran University of Medical Sciences; [4]Department of Pulmonary Medicine, Shariati Hospital, Tehran University of Medical Sciences; [5]Department of Radiology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

Abstract

Objective. Systemic sclerosis-associated interstitial lung disease (SSc-ILD) is a major cause of morbidity and mortality in systemic sclerosis patients. High-resolution computed tomography (HRCT) plays a crucial role in SSc-ILD diagnosis and management. In this study, we aimed to develop and evaluate a deep learning-based automated segmentation model for quantifying SSc-ILD lesions in HRCT images and assess its clinical relevance.
Methods. We developed a convolutional neural network model to segment normal lung, established fibrosis (EF), and ground-glass opacity (GGO) in HRCT scans from 40 SSc-ILD patients. The model was trained and evaluated using 8-fold cross-validation. Segmentation performance was assessed using the Dice similarity coefficient (DSC). Correlations between predicted lesion volumes and pulmonary function test (PFT) metrics were analyzed using Spearman's ρ.
Results. Our model achieved a total lesion (EF+GGO) DSC of 78%. Class-wise segmentation performance was lower for EF (DSC: 70%) compared to GGO (DSC: 73%). Predicted lesion volumes showed significant negative correlations with forced expiratory volume in one second (FEV_1) (ρ=-0.64, $p<0.001$) and FEV1/forced vital capacity (ρ=-0.73, $p<0.001$). We also created the SICCS dataset, a public dataset of SSc-ILD HRCT images with expert-annotated segmentation labels.
Conclusions. Deep learning-based automated segmentation can help quantify SSc-ILD lesions in HRCT images and provide clinically relevant information. The model's performance is comparable to previous studies, and the predicted lesion volumes correlate significantly with PFT metrics. This approach shows promise for aiding in SSc-ILD diagnosis, monitoring, and clinical decision-making, although further validation with larger datasets is needed.

Key words: systemic sclerosis, interstitial lung disease, SSc-ILD, segmentation, HRCT, deep learning.

Correspondence: Hoda Kavosi, Rheumatology Research Center, Shariati Hospital, Kargar Ave, Tehran, 14117-13137, Iran.
E-mail: h-kavosi@sina.tums.ac.ir

Introduction

Systemic sclerosis (SSc) is a systemic connective tissue disorder defined by the development of fibrosis and atrophy in the skin, subcutaneous tissue, muscle tissue, and internal organs. SSc primarily affects women, with the usual age of onset ranging from 45 to 64 years (1, 2). This condition is associated with considerable morbidity and mortality, with pulmonary complications as the primary cause of SSc-related case fatalities. Interstitial lung disease due to SSc (SSc-ILD), a common pulmonary consequence of SSc, affects a significant proportion of SSc patients (70-90%) and is responsible for the majority of SSc-related deaths (up to 17%) (3, 4).

SSc-ILD is characterized by inflammation and fibrosis of the lung parenchyma, leading to progressive decline in lung function and reduced quality of life. The course of the disease displays considerable variation, ranging from subclinical lung involvement to major pulmonary disease, respiratory failure, and death. While certain patients may stay steady for an extended period, others undergo phases of deterioration succeeded by stability or ongoing progression (5, 6).

High-resolution computed tomography (HRCT) of the chest is considered the gold standard for detecting SSc-ILD (7), capable of detecting pulmonary changes even in patients with no pulmonary symptoms and/or normal pulmonary function test (PFT) results (8, 9). The importance of HRCT in SSc-ILD management extends beyond initial diagnosis. HRCT is valuable for disease monitoring, determining prognosis, and potentially predicting treatment response. The extent of ILD involvement in HRCT has been shown to correlate with functional impairment, pulmonary function decline, and mortality risk (6, 10). From a radiological perspective, the common progression pattern in SSc-ILD is nonspecific interstitial pneumonia, dominated by ground-glass opacity (GGO) patterns, followed by usual interstitial pneumonia, manifesting as honeycombing and traction bronchiectasis. Despite these advan-

tages, radiation exposure remains an ongoing consideration, and routine baseline HRCT for SSc patients has not been recommended by expert groups (4).

Given the importance of HRCT in SSc-ILD management, accurate quantification of disease extent and progression is essential. Automated segmentation and volume prediction of SSc-ILD lesions in HRCT images offer several potential benefits, including objective, reproducible measurements of disease extent, allowing for more precise monitoring of disease progression and treatment response.

Building on the proven track record of deep learning models for medical image segmentation tasks (11-13), we developed an automated SSc-ILD lung lesion segmentation model, which, to the best of our knowledge, is the first multiclass segmentation model available for this task. We also evaluate the relation between the lesion volume predicted by our model and routine PFT metrics used for monitoring SSc-ILD. Finally, we have also made our dataset public under the title of SICCS (SSc-ILD chest CT segmentation) (14), so future research by other teams can also use our data, including images and segmentation labels.

Key messages

What is already known on this topic: SSc-ILD is a leading cause of morbidity and mortality in SSc, and while HRCT is the gold standard for detecting lung involvement, quantification of fibrosis and GGO remains manual, time-consuming, and lacks robust multiclass automated methods. What this study adds: we developed and validated a ConvNext V2-inspired U-Net model, achieving a total Dice similarity coefficient (DSC) of 0.78 for established fibrosis (EF) and GGO segmentation in SSc-ILD; demonstrated strong inverse correlations between predicted lesion volumes and key PFT metrics [forced expiratory volume in one second (FEV_1), FEV_1/forced vital capacity (FVC)]; and released SICCS, the first public multiclass SSc-ILD HRCT segmentation dataset. How this study might affect research, practice, or policy: our automated, reproducible quantification tool has the potential to streamline clinical monitoring and treatment evaluation in SSc-ILD, and the publicly available SICCS dataset will catalyze further artificial intelligence-driven research and algorithm development in this field.

Methods

Data

Chest CT images, PFT data, and demographic data were retrospectively collected for patients who met the American College of Rheumatology/European Alliance of Associations for Rheumatology criteria for SSc (15), and were undergoing treatment for SSc at Shariati Hospital, Tehran, Iran, from electronic records (data accessed in November 2024). Originally, these were recorded during 2019-2023, and all CT images were acquired a maximum of 15 days before or after the PFT records. Written informed consent regarding anonymized use of data for research purposes was acquired from all participants for both hospital data and image data. Chest CT images for the study population comprised a heterogeneous dataset acquired from multiple imaging centers with different resolutions and slice thickness values (0.7-5mm). While challenging in terms of segmentation performance, this is also an advantage since it better reflects actual clinical use case scenarios with less-than-ideal data.

Preprocessing

Preprocessing for our publicly available data consisted of conversion from DICOM to NIfTI file format and de-identification of all the header data in the files (16). For our model, we resized all the images to a resolution of 320×320×40 voxels, *z*-score normalized each fold of the data (cross-validation schema explained below) according to the training data average and standard deviation, resampled all scans to an isotropic spacing of 1 mm^3, and performed histogram equalization using the CLAHE algorithm to mitigate the effects of data heterogeneity (17, 18).

Segmentation ground-truth acquisition

The images were labeled based on consensus from a pulmonologist (MA, the 8th author of this paper) and a radiologist (LA, the 9th author of this paper), with at least 7 years of experience in using HRCT to manage SSc-ILD. CT images were labeled according to the following schema: 0) background; 1) normal lung; 2) EF, referring to late-stage radiological signs such as honeycombing; 3) GGO. Labeling was performed in 3D Slicer (19), using axial plane views and then revised for accurate contours in the coronal and sagittal planes. We used the DSC (explained later) as a measure of interrater reliability, with an average DSC for the lesions at 84% and for the normal lung at 98% when comparing the two label sets prepared. Discrepancies were reviewed in joint consensus sessions using multi-planar views until agreement was achieved. All annotations followed a standardized protocol based on Fleischner Society guidelines for HRCT pattern interpretation, ensuring consistent delineation of GGO and fibrosis regions (20). The final label masks were visually inspected by both experts prior to inclusion in the public SICCS dataset.

Segmentation model

We used a U-net variant inspired by the ConvNext V2 family of models (21, 22). Briefly, our model has a convolutional encoder-decoder architecture with skip connections, processing the image at four resolutions. We used an input layer with a resolution of 320×320×40, halving the dimensions at each level, with the lowest size being 40×40×5. Both the encoder and decoder used 1, 1, 3, and 1 block for stages 1, 2, 3, and 4, respectively. This architecture was designed to leverage specific techniques from the ConvNext architecture, such as Global Response Normalization.

Training strategy

We trained and evaluated our model using 8-fold cross-validation, with a 35-5 split schema. Data augmentation was used to address the low sample size, using random intensity scale, random Gaussian noise, random contrast adjustment, and random Gaussian smoothing with reasonable parameters; this augmentation strategy also improves robustness to variations in intensity distributions and acquisition noise. To quantify the impact of heterogeneity, we performed an exploratory analysis by grouping scans by slice thickness (<1 mm *vs.* ≥1 mm) and confirmed that no significant difference in DSC was observed between groups ($p>0.05$). This suggests that our preprocessing and augmentation strategies effectively mitigated scanner-related bias. For optimization, we used the Focal Dice loss function, which is a weighted sum of the focal loss (23) and the dice loss (24) functions [Eq. 1]:

$$L_{\text{focal Dice}} = \lambda_{\text{focal}} \cdot -(1-p_{\text{pred}})^{\gamma} \cdot \log(p_{\text{pred}}) + \lambda_{\text{Dice}} \cdot \left(1 - \frac{2\times\sum p_{\text{true}}\times p_{\text{pred}}}{\sum p_{\text{true}}^2 + \sum p_{\text{pred}}^2 + \varepsilon}\right)$$

[Eq. 1]

We trained all 8 models for a fixed number of epochs (100), using a cosine annealing learning rate schedule and the AdamW optimizer with AMSGrad (25, 26). Since we solely used cross-validation for our evaluations, we refrained from hyperparameter tuning. Detailed hyperparameter configuration for model training is available in *Supplementary Table 1*.

Statistical analysis

We used Spearman's ρ and p-values to assess the relationship between predicted lesion volume (% of total lung volume) and ground truth lesion volume, FEV_1/FVC, and FEV_1 (% predicted), based on the validation folds of our cross-validation schema. Descriptive statistics are reported using percentages (for categorical variables) and mean±standard error values (for continuous variables).

Segmentation performance was evaluated using the DSC formula, which measures the overlap between the predicted and ground truth masks [Eq. 2]:

$$\text{Dice coefficient} = \frac{2\times TP}{TP+TN+FP+FN} \quad \text{[Eq. 2]}$$

where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively.

Software

All the stages of the analysis for this work were performed using the Python programming language (version 3.10), utilizing the following libraries: i) PyTorch v2.1 for model design and training (27); ii) MONAI for preprocessing and data augmentation (28); iii) Scikit-image for the CLAHE algorithm (29); iv) SciPy stats module for statistical analysis (30).

Patient and public involvement

No patients or members of the public were involved in the design, conduct, choice of outcome measures, recruitment, or dissemination plans for this study.

Results

40 SSc-ILD patients were included in the study. The mean age was 49.3 years (standard deviation 14.29), and only 10% of the study population were male. Average FEV_1/FVC ratio was 86.7%, and mean FEV_1 (% predicted) was 73.2%. More details, demographic and PFT-related values are available in *Supplementary Table 2*.

Robust multi-class segmentation performance

For segmentation performance, we evaluated our model using both DSC and lesion volume prediction accuracy. Across all eight cross-validation folds, our model achieved a mean DSC of 78.2% for combined lesions (EF+GGO). Class-wise DSC values were 87.4% for normal lung tissue, 70.1% for EF, and 73.1% for GGO (Table 1). We have also provided visual examples in Figure 1. This segmentation performance is a key factor for our downstream analyses on lesion volume prediction and PFT metrics.

Accurate lesion volume prediction

Predicted lesion volumes were highly correlated to ground truth volumes (EF, GGO, and aggregated EF+GGO all had $p<0.001$). GGO and EF-GGO had Spearman's ρ values of 0.97 and 0.98, respectively; volume prediction for EF alone was comparatively less accurate ($\rho=0.86$). A subject-level assessment of lesion volume prediction accuracy is available in Figure 2.

Lesion volume links to functional lung impairment

To assess the clinical utility of our model, we correlated total lesion burden (volume) with PFT metrics. Higher lesion burden was significantly associated with deteriorating lung function: Spearman's ρ values for predicted lesion burden (EF+GGO volume) *vs*. FEV_1/FVC and FEV_1 (% predicted) were -0.64 and -0.73, respectively, and both correlations were significant ($p<0.001$). These clinical implications of our model's predictions are also elaborated in Figure 3 and Table 2.

Discussion

In this study, we developed a deep learning model for automated segmentation of SSc-ILD lesions on HRCT, achieving a mean total lesion DSC of 78%, which is higher than the 71-75% reported in the only comparable study using deep learning for SSc-ILD segmentation (31). This improvement likely reflects our use of a ConvNext-inspired U-Net architecture and extensive data augmentation, both of which enhance generalization to heterogeneous clinical scans. Comparable DSC values have been reported in pul-

Table 1. Segmentation performance.

Class	DSC	ρ (predicted *vs.* ground truth volume)	p
Normal Lung	87.4±2.3%	-	-
EF	70.1%±4.4%	0.867	<0.001
GGO	73.1%±4.3%	0.979	<0.001
EF+GGO	78.2±3.2%	0.970	<0.001

DSC, Dice similarity coefficient; EF, established fibrosis; GGO, ground-glass opacity. DSC values are reported as mean ± standard error values.

Table 2. Predicted lesion volume *vs.* pulmonary function test metrics.

PFT metric	ρ (predicted volume *vs.* metric)	p
FEV_1/FVC	-0.649	<0.001
FEV_1	-0.736	<0.001

PFT, pulmonary function test; FEV1, forced expiratory volume in one second; FVC, forced vital capacity.

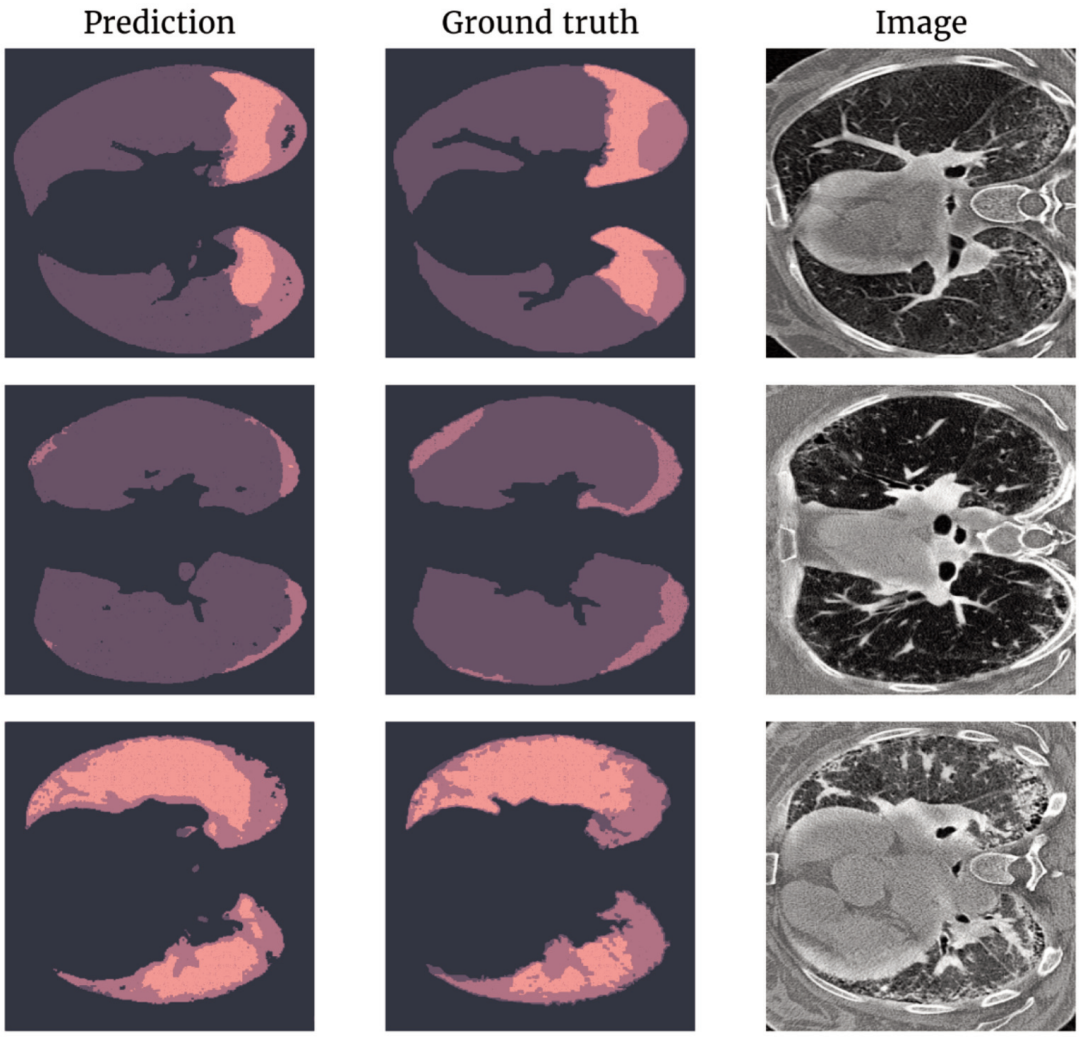

Figure 1. Representative high-resolution computed tomography slices with predicted (left) and ground-truth (middle) segmentations, showing various extents of lesion volume in the dataset. Purple: normal lung; pink: established fibrosis; orange: ground-glass opacity.

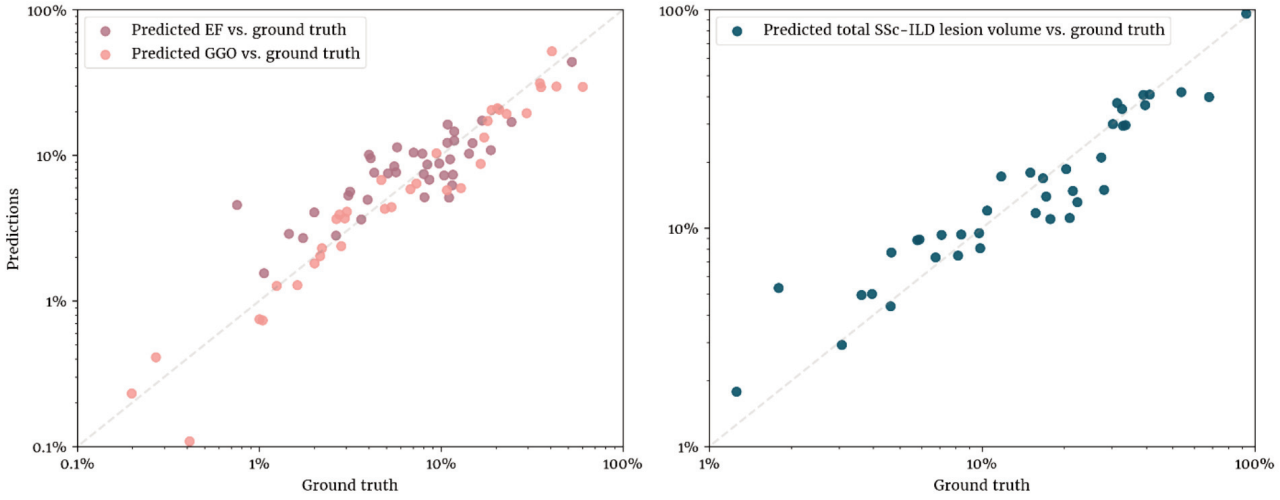

Figure 2. Log-log scatter plot of predicted *vs.* ground−truth lesion volumes [left: class-wise, right: aggregated established fibrosis (EF) + ground-glass opacity (GGO)] across all 40 subjects in cross-validation. The dashed line is the identity line (y = x), indicating ideal correlation. SSc-ILD, systemic sclerosis-associated interstitial lung disease.

monary lesion segmentation studies for other etiologies, such as ILD and COVID-19 (32, 33), supporting the robustness of our approach.

Class-wise segmentation performance was not in keeping with our initial expectations (better performance in EF segmentation compared to GGO). We hypothesize several factors could have contributed to this observation, such as the relatively smaller portion of EF available in our dataset (~8% based on volume) compared to GGO (~11%), and also the limitation of the DSC in evaluating segmentation performance for structures with irregular contours. Moreover, the EF pattern in CT imaging is inherently heterogeneous according to radiology literature (34-36), which further complicates the task of developing automated segmentation solutions for this pattern of pulmonary lesions.

From a clinical standpoint, the relatively lower DSC for EF is important to consider because fibrotic changes on HRCT and reductions in DLCO are among the earliest and most prognostically relevant indicators of SSc-ILD progression. In clinical practice, EF extent and DLCO decline are often used together to stratify disease severity and predict outcomes (4). The reduced EF segmentation accuracy in our model, therefore, suggests that quantitative estimates of fibrotic burden should be interpreted cautiously, particularly for early-stage disease or longitudinal follow-up where subtle fibrotic progression is clinically meaningful. Nonetheless, the strong correlation between total lesion volume (EF + GGO) and pulmonary function parameters (FEV_1, FEV_1/FVC) indicates that our model still captures global structural and functional deterioration effectively.

Our study also assessed the utility of our predicted lesion volume in correlating with PFT metrics. In contrast to semantic segmentation, which is underexplored, there are several studies available in the literature applying quantitative analysis techniques to HRCT images aiming to correlate the outputs to PFT metrics (31, 37-41), including some studies focusing on idiopathic pulmonary fibrosis unrelated to SSc-ILD; nevertheless, our findings were congruent with the available evidence, indicating statistically significant correlations between HRCT findings and PFT metrics.

One notable quantitative approach to lung HRCT assessment is the quantitative lung fibrosis (QLF) tool (42), which is comparable to our approach in the sense that both methods can be corroborated against PFT metrics. According to a systematic review in 2025, in the context of ILD, QLF findings are also correlated to pulmonary functions, with weaker correlations compared to our results (such as r=-0.33 for QLF score *vs.* FVC) (43). Interestingly, one study points to possibly distinct pathological processes for GGO and QLF findings (which focuses on fibrosis) in the context of SSc-ILD, as evident in the weak correlation reported by the authors (44). More comprehensive studies will be needed to accurately compare the various approaches to lung HRCT assessment for SSc-ILD. A final important aspect of our work is our public dataset (14), which is beneficial to research efforts focused on SSc-ILD as a whole. This is evident in the case of well-established public segmentation datasets for other tasks, such as glioma segmentation in MRI and ultrasound (45, 46), and pediatric bone age regression based on hand radiographs (47).

Limitations

There are several limitations in our study, including a small sample size (and lack of a proper hold-out test set), which limits the robustness and generalizability of our segmentation model. Nonetheless, since no hyperparameter tuning was performed in this work, cross-validation is equivalent to using several hold-out test sets. The heterogeneity of our dataset in terms of imaging characteristics, such as image acquisition device, slice thickness, resolution, *etc*., is unconventional compared to most other publicly available datasets for medical image processing research, but also provides an opportunity in terms of dealing with data more similar to real-world scenarios, which might not come from a single imaging center. Our options in terms of PFT data were also limited: while most of the studies mentioned earlier also used diffusing capacity of the lung for carbon monoxide (DLCO), this test was only available for 8 of our subjects (out of n=40), and thus we decided not to use DLCO in our analysis; DLCO testing is not routinely available for all patients in our setting, as it requires specialized equipment and calibration that are not consistently accessible in public-sector hospitals in Iran. One final limitation is our deci-

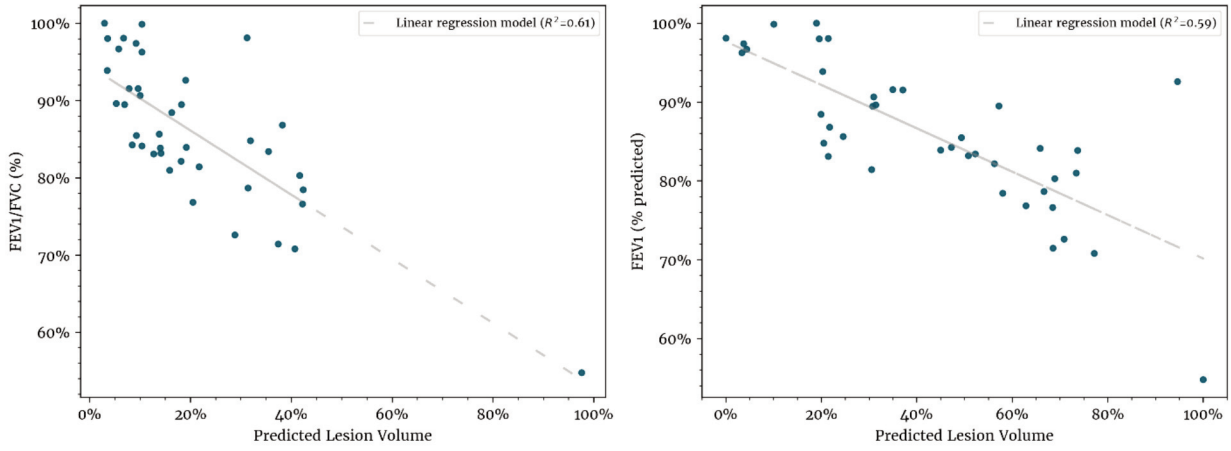

Figure 3. Scatter plot of aggregate established fibrosis + ground-glass opacity predicted lesion volume *vs.* forced expiratory volume in one second (FEV_1)/forced vital capacity (FVC) (left) and FEV_1 (right). The dashed line represents fitted linear regression models to the data, for illustration purposes.

sion not to utilize other radiologic findings attributed to SSc-ILD in HRCT images for our model, such as traction bronchiectasis (48). Our focus on GGO and EF as the two main classes of pulmonary lesions associated with SSc-ILD stemmed from the evidence pointing to more pronounced roles for these two classes of lesions in SSc-ILD prognosis (49), and also the smaller prevalence of other lesion classes in our data.

Future work

In future work, we plan expand upon our preliminary dataset with an increased number of subjects, temporal follow-up data, and a more comprehensive approach to PFT metrics. In particular, larger, multi-center datasets will enable improved evaluation schemas with proper test sets and external validation. Moreover, with the increasing focus on utilizing multimodal data for medical predictive models (50), ultrasound imaging of the lung is another direction we aim to pursue. While not yet established in clinical practice due to several factors, including lack of established protocols, this modality also has distinct advantages, such as being ionizing radiation-free (49, 51); combined with computer-assisted quantitative techniques such as automated segmentation, we believe multimodal approaches have significant promise in improving care standards for SSc-ILD patients.

Conclusions

In this work, we have developed a deep learning model for automated HRCT pulmonary lesion segmentation in a population of SSc-ILD patients, with performance on par with the current state-of-the-art in terms of segmentation and correlation with PFT metrics. If integrated into clinical frameworks, our model has important uses in automating and streamlining SSc-ILD diagnosis and assisting clinical experts with reliable quantitative assessments of ILD extent. We have also contributed a publicly available segmentation dataset to the field of SSc-ILD research, which we believe can be immensely valuable in future research based on how beneficial other datasets have been to research on medical image processing.

References

1. Moghadam KG, Gharibdoost F, Parastandechehr G, Salehian P. Assessments of pulmonary involvement in patients with systemic sclerosis. Arch Iran Med 2011; 14: 22-6.
2. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Semin Arthritis Rheum 2008; 37: 223-35.
3. Elhai M, Meune C, Boubaya M, Avouac J, Hachulla E, Balbir-Gurman A, et al. Mapping and predicting mortality from systemic sclerosis. Ann Rheum Dis 2017; 76: 1897-905.
4. Ruaro B, Baratella E, Confalonieri P, Wade B, Marrocchio C, Geri P, et al. High-resolution computed tomography: lights and shadows in improving care for SSc-ILD patients. Diagnostics 2021; 11: 1960.
5. Hoffmann-Vold AM, Allanore Y, Alves M, Brunborg C, Airó P, Ananieva LP, et al. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. Ann Rheum Dis 2021; 80: 219-27.
6. Moore OA, Goh N, Corte T, Rouse H, Hennessy O, Thakkar V, 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 2013; 52: 155-60.
7. Hoffmann-Vold AM, Maher T, Philpot EE, Ashrafzadeh A, Barake R, Barsotti S, et al. The identification and management of interstitial lung disease in systemic sclerosis: evidence-based European consensus statements. Lancet Rheumatol 2020; 2: e71-83.
8. Suliman YA, Dobrota R, Huscher D, Nguyen-Kim TDL, Maurer B, Jordan S, 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: 3256-61.
9. Showalter K, Hoffmann A, Rouleau G, Aaby D, Lee J, Richardson C, et al. Performance of forced vital capacity and lung diffusion cutpoints for associated radiographic interstitial lung disease in systemic sclerosis. J Rheumatol 2018; 45: 1572-6.
10. Khanna D, Nagaraja V, Tseng CH, Abtin F, Suh R, Kim G, et al. Predictors of lung function decline in scleroderma-related interstitial lung disease based on high-resolution computed tomography: implications for cohort enrichment in systemic sclerosis-associated interstitial lung disease trials. Arthritis Res Ther 2015; 17: 372.
11. Cui H, Yuwen C, Jiang L, Xia Y, Zhang Y. Multiscale attention guided U-Net architecture for cardiac segmentation in short-axis MRI images. Comput Methods Programs Biomed 2021; 206: 106142.
12. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. Comput Methods Programs Biomed 2018; 161: 1-13.
13. Li C, Song X, Zhao H, Feng L, Hu T, Zhang Y, et al. An 8-layer residual U-Net with deep supervision for segmentation of the left ventricle in cardiac CT angiography. Comput Methods Programs Biomed 2021; 200: 105876.
14. Vahdani A. SICCS: SSc-ILD chest CT segmentation. Available from: https://doi.org/10.7910/DVN/3HNP33.
15. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum 2013; 65: 2737-47.
16. National Institutes of Health. NIfTI: neuroimaging informatics technology initiative. Available from: https://nifti.nimh.nih.gov/.
17. Pizer SM, Johnston RE, Ericksen JP, Yankaskas BC, Muller KE. Contrast-limited adaptive histogram equalization: speed and effectiveness. First Conference on Visualization in Biomedical Computing, pp. 337-45.
18. Salem N, Malik H, Shams A. Medical image enhancement based on histogram algorithms. Procedia Comput Sci 2019; 163: 300-11.
19. Kikinis R, Pieper SD, Vosburgh KG. 3D slicer: a platform for subject-specific image analysis, visualization, and clinical support. In: Jolesz FA, ed. Intraoperative imaging and image-guided therapy. New York, NY: Springer New York; 2014. pp. 277-89.
20. Bankier AA, MacMahon H, Colby T, Gevenois PA, Goo JM, Leung ANC, et al. Fleischner Society: glossary of terms for thoracic imaging. Radiology 2024; 310: e232558.

21. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham; 2015. pp. 234-41.
22. Woo S, Debnath S, Hu R, Chen X, Liu Z, Kweon IS, et al. Convnext v2: co-designing and scaling convnets with masked autoencoders. IN: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16133-42.
23. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.
24. Milletari F, Ahmadi SA, Kroll C, Plate A, Rozanski V, Maiostre J, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Comput Vis Image Underst 2017; 164: 92-102.
25. Loshchilov I, Hutter F. Fixing weight decay regularization in adam. ArXiv 2017; abs/1711.05101.
26. Reddi SJ, Kale S, Kumar S. On the convergence of adam and beyond. arXiv; 2019: 1904.09237.
27. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst 2019; 32.
28. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, et al. Monai: an open-source framework for deep learning in healthcare. arXiv 2022; 2211.02701.
29. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. Scikit-image: image processing in Python. PeerJ 2014; 2: e453.
30. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020; 17: 261-72.
31. Chassagnon G, Vakalopoulou M, Régent A, Zacharaki EI, Aviram G, Martin C, et al. Deep Learning-based Approach for Automated Assessment of Interstitial Lung Disease in Systemic Sclerosis on CT images. Radiol Artif Intell 2020; 2: e190006.
32. Hoang-Thi TN, Vakalopoulou M, Christodoulidis S, Paragios N, Reve MPl, Chassagnon G. Deep learning for lung disease segmentation on CT: Which reconstruction kernel should be used? Diagn Interv Imaging 2021; 102: 691-5.
33. Khomduean P, Phuaudomcharoen P, Boonchu T, Taetragool U, Chamchoy K, Wimolsiri N, et al. Segmentation of lung lobes and lesions in chest CT for the classification of COVID-19 severity. Sci Rep 2023; 13: 20899.
34. Brady D, Berkowitz EA, Sharma A, Ackman JB, Bernheim A, Chung M, et al. CT morphologic characteristics and variant patterns of interstitial pulmonary fibrosis in systemic lupus erythematosus. Radiol Cardiothorac Imaging 2021; 3: e200625.
35. Gruden JF. CT in idiopathic pulmonary fibrosis: diagnosis and beyond. Am J Roentgenol 2016; 206: 495-507.
36. Park J, Jung J, Yoon SH, Hong H, Kim H, Kim H, et al. CT quantification of the heterogeneity of fibrosis boundaries in idiopathic pulmonary fibrosis. Eur Radiol 2021; 31: 5148-59.
37. Humphries SM, Yagihashi K, Huckleberry J, Rho BH, Schroeder JD, Strand M, et al. Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up. Radiology 2017; 285: 270-8.
38. Humphries SM, Swigris JJ, Brown KK, Strand M, Gong Q, Sundy JS, et al. Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. Eur Respir J 2018; 52: 1801384.
39. Carvalho ARS, Guimarães AR, Sztajnbok FR, Souza Rodrigues R, Antunes Silva BR, Lopes AJ, et al. Automatic quantification of interstitial lung disease from chest computed tomography in systemic sclerosis. Front Med 2020; 7: 577739.
40. Tashkin DP, Volkmann ER, Tseng CH, Kim HJ, Goldin J, Clements P, et al. Relationship between quantitative radiographic assessments of interstitial lung disease and physiological and clinical features of systemic sclerosis. Ann Rheum Dis 2016; 75: 374-81.
41. Ooi GC, Mok MY, Tsang KWT, WOng Y, Khong PL, Fung PCW, et al. Interstitial lung disease in systemic sclerosis. Acta Radiol 2003; 44: 258-64.
42. Kim HG, Tashkin DP, Clements PJ, Li G, Brown MS, Elashoff R, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 2010; 28: S26-35.
43. Dixon G, Thould H, Wells M, Tsaneva-Atanasova K, Scotton CJ, Gibbons MA, et al. A systematic review of the role of quantitative CT in the prognostication and disease monitoring of interstitial lung disease. Eur Respir Rev 2025; 34: 240194.
44. Volkmann ER, Tashkin DP, Leng M, Kim GHJ, Goldin J, Harui A, Roth MD. Biological correlates of radiological features of systemic sclerosis interstitial lung disease. ERJ Open Res 2025; 11: 00596-2024.
45. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J. The multimodal brain tumor image segmentation benchmark. IEEE Trans Med Iimaging 2015; 34: 1993-2024.
46. Juvekar P, Dorent R, Kögl F, Torio E, Barr C, Rigolo L, et al. ReMIND: the brain resection multimodal imaging database. medRxiv 2024; 2023.09.14.23295596.
47. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, et al. The RSNA Pediatric Bone Age Machine Learning Challenge. Radiology 2018; 290: 498-503.
48. de Lima Bastos A, de Amorim Corrêa R, Ferreira GA. Tomography patterns of lung disease in systemic sclerosis. Radiol Bras 2016; 49: 316-21.
49. Landini N, Orlandi M, Bruni C, Carlesi E, Nardi C, Calistri L, et al. Computed tomography predictors of mortality or disease progression in systemic sclerosis-interstitial lung disease: a systematic review. Front Med 2022; 8: 807982.
50. Simon BD, Ozyoruk KB, Gelikman DG, Harmon SA, Türkbey B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. Diagn Interv Radiol 2025; 31: 303-12.
51. Hughes M, Bruni C, Cuomo G, Delle Sedie A, Gargani L, Gutierrez M, et al. The role of ultrasound in systemic sclerosis: On the cutting edge to foster clinical and research advancement. J Scleroderma Relat Disord 2021; 6: 123-32.

Online supplementary material:
Supplementary Table 1. Training configuration.
Supplementary Table 2. Baseline data of the study population (n=40).

Received: 28 June 2025; Accepted: 8 January 2026; Early view: 25 February 2026.

Contributions: Amir M. Vahdani: methodology, software, writing – original draft, formal analysis, visualization. Aida Mohamadi: investigation, data curation, writing – review and editing. Sepehr Nayebirad: conceptualization, methodology, software, formal analysis, writing – review and editing. Shayan Forghani: investigation, data curation, writing – review and editing. Elaheh Karimi: investigation, data curation, writing – review and editing. Yousef Vojgani: investigation, data curation, writing – review and editing. Zahra Tamartash: investigation, data curation, writing, review, and editing. Mehrnaz Asadi: investigation, data curation, writing – review and editing. Leila Aghaghazvini: investigation, data curation, writing – review and editing. Farhad Gharibdoost: investigation, data curation, writing – review and editing. Hoda Kavosi: conceptualization, writing – review and editing, supervision, project administration.

Conflict of interest: the authors declare that they have no conflicts of interest. There are no financial or non-financial interests that could directly or indirectly influence the work reported in this study.

Ethics approval and consent to participate: this study was conducted in accordance with ethical standards and relevant institutional guidelines. All procedures involving human participants were reviewed and approved by the Ethics Committee of Tehran University of Medical Sciences (approval code: IR.TUMS.MEDICINE.REC.1403.297) on 10/02/2024.

Informed consent: written informed consent was obtained from all individual participants included in the study prior to their enrollment. Participants were informed about the purpose of the research, the procedures involved, and their right to withdraw at any time without consequence.

Patient consent for publication: the authors certify that they have obtained all appropriate patient consent forms. In these forms, patients have given their consent for the anonymized use of their clinical and imaging data for research purposes, as well as the publishing of their chest CT images (which contain no sort of identifying information) as a publicized dataset.

Availability of data and materials: the complete dataset used in this work—consisting of de‑identified SSc‑ILD HRCT scans and expert‑annotated segmentation labels—is publicly accessible at https://doi.org/10.7910/DVN/3HNP33.