BMC Pulmonary Medicine

RESEARCH

Open Access

A noval identification of 4 systemic sclerosis - interstitial lung disease subgroups using principal component analysis-based cluster analysis

Yaqi Zhao[2†], Baoting Chao[4†], Wei Xu[2,3], Xinya Li[3], Jin Zhang[3], Ying Zhang[3], Zhenzhen Ma[1,2*] and Qingrui Yang[2,3*]

Abstract

Objective Interstitial lung disease (ILD) is a common and serious complication of systemic sclerosis (SSc). It is usually classified by histologic type, but this classification may not fully reflect the clinical phenotypic variation. This study aimed to examine clinical features and aggregate patients with SSc-ILD based on patients' clinical manifestations, High-resolution computed tomography (HRCT) features and specific antibody expression to achieve precise treatment of SSc-ILD with early identification of complications.

Methods This study included 103 patients with SSc-ILD. A cluster analysis was performed based on five clinical and serological variables to identify subgroups of patients. The survival rates between obtained clusters and risk factors affecting prognosis were also compared.

Result Four clusters were identified in this study: Cluster 1 ($n=23$) represented the lymphocytic interstitial pneumonia (LIP) group with LIP as the predominant HRCT characteristic. Cluster 2 ($n=23$) was the worst prognosis group, with the highest Warrick score as well as the highest mortality rate. Cluster 3 ($n=20$) with all patients having a negative anti-SCL-70 antibody response. Cluster 4 ($n=28$) with all patients were positive for the anti-SCL-70 antibody. It was found that albumin was a protective factor for the prognosis of patients with SSC-ILD patients ($p=0.018$), whereas age ($p=0.036$) and IgM ($p=0.040$) were risk factors.

Conclusion The results of our cluster analysis indicated that based solely on histologic typing, may not be capturing the full heterogeneity of SSc-ILD patients. In order to identify homogeneous patient groups with a specific prognosis, HRCT features and antibody profiles should be taken into consideration.

†Yaqi Zhao and Baoting Chao contributed equally to this work and share first authorship.

*Correspondence:
Zhenzhen Ma
mazhenzhendz@163.com
Qingrui Yang
qryang720@163.com

Full list of author information is available at the end of the article

Significance and innovations

This study examined clinical features and aggregate patients with SSc-ILD based on patients' clinical manifestations, High-resolution computed tomography (HRCT) features and specific antibody expression to achieve precise treatment of SSc-ILD with early identification of complications.

Keywords Systemic sclerosis, Interstitial lung disease, Cluster analysis, HRCT, Warrick score

Introduction

Systemic sclerosis (SSc) is a systemic autoimmune disease characterized by microangiopathy with progressive loss of capillaries, mild inflammation and an overproduction of extracellular matrix proteins by fibroblasts [1]. Clinical characteristics of SSc include skin fibrosis in various degrees and internal organ involvement, pulmonary involvement is the second most frequent visceral complication of SSc [2] and the leading cause of disease-related death [3].

Despite the availability of multiple controlled clinical trials, international guidelines, and consensus documents regarding immunosuppressive and antifibrotic therapies for the treatment of SSc-ILD, there remains a critical need for personalized therapeutic strategies [4]. Additionally, hastily seeking medical attention could expose stable patients to needless drug toxicants and increase their risk of infection [5]. Although pulmonary function tests, clinical variables, and high-resolution computed tomography (HRCT)-based assessments of disease extent are widely recognized as established prognostic markers for SSc-ILD, existing classification systems often fall short in capturing the inherent heterogeneity of this disease. In clinical practice, surgical lung biopsy, which has historically been regarded as the gold standard for ILD classification, is rarely performed in patients with SSc-ILD due to the associated risks of complications and concerns regarding sampling variability. High-resolution computed tomography (HRCT) [6] has been proven to allow early detection of lung involvement, even at subclinical stages, and to assess ILD both qualitatively (by reviewing the pattern of interstitial pneumonia) and quantitatively (by determining the extent of ILD) which has become the gold standard for the detection of ILD [7]. Given the similarities in pathological patterns and the role of staging in guiding clinical treatment, SSc-ILD can be managed according to the HRCT staging criteria used for idiopathic interstitial pneumonias (IIP) [8]. HRCT is a valuable tool for identifying predominant interstitial pneumonia patterns in SSc-ILD, such as usual interstitial pneumonia (UIP), nonspecific interstitial pneumonia (NSIP), organizing pneumonia (OP), and lymphocytic interstitial pneumonia (LIP). However, it may not reliably differentiate all subtypes (e.g., lymphocytic variants) without additional clinical or histologic data. According to a study, some fibrotic ILD patients who may have a progressive phenotype exhibit a UIP pattern. In patients with rheumatoid arthritis (RA)-associated interstitial lung disease (ILD), those presenting with a usual interstitial pneumonia (UIP) pattern on HRCT have a significantly worse prognosis than those without this pattern. This underscores the potential importance of identifying this specific imaging feature in clinical assessment [9].

Unlike previous studies, not only is the study based on clinical data, it is a joint study that combines imaging information with clinical data. We used HRCT to stage SSc-ILD patients and quantified the degree of pulmonary fibrosis to develop a clinical prediction model for patients with severe ILD. And using cluster analysis to detect and characterize homogeneous groups of SSc-ILD patients as well as to compare survival rates and factors affecting survival prognosis between clusters, in order to achieve precise treatment of SSc early prevention of complications.

Materials and methods

Study design and population

This study included SSc-ILD patients diagnosed during hospitalization from 2013 to 2022 (All patients were from Shandong Provincial Hospital). Patients were included when they satisfied all of the following conditions: (1) met the 2013 ACR/EULAR SSc classification criteria [10] and was diagnosed by two or more rheumatologists, (2) ILD was defined by the following criteria at any point during the patient's clinical course: (1) Any of the following chest HRCT findings and exclusion of patients with respiratory infections: (a) Ground-glass/reticular opacities; (b) Bronchiectasis; (c) Interlobular septal/subpleural thickening; (d)Honeycombing; (e) Consolidations or cysts. (2) Confirmation of ILD diagnosis by primary rheumatologist/pulmonologist [11], and (3) underwent pulmonary function tests (PFTs) and laboratory examination within 30 days before or after the CT scan. Patients with other rheumatic diseases and SSc patients combined with Pulmonary hypertension (PH) were excluded from this study. In order to minimize the influence of overlapping syndromes on the experimental results and to avoid the inclusion of pathological processes as well as clinical manifestations of PH in the study. Specific study population choices are described in Supplementary material 1. Chest tightness is defined as the patient's subjective feeling of labored breathing or lack of breath, which may be accompanied by shortness of breath. Smoke was defined

as being active smoking and having quit smoking for less than 12 months. Esophageal involvement includes Gastroesophageal Reflux Disease (GERD)/Reflux Esophagitis (RE). Cutaneous involvement is defined as digital swelling (puffy fingers); skin thickening, hardening and tightening; acrolysis, flexion contracture, salt and pepper skin [12]. We defined systemic sclerosis renal crisis (SRC) as the combination of systemic hypertension ≥ 150/85 mmHg and a decrease in estimated glomerular filtration rate ≥ 30% or features in renal biopsy [13]. Cardiac involvement defined as an increase in NT-proBNP plus cardiac troponin T and/or structural and/or functional damage diagnosed on echocardiography/cardiac MRI that showed left ventricular dysfunction in absence of better causes that explain the features [14]. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Shandong Provincial Hospital, Shandong First Medical University (NSFC: NO.2022 – 413). All participants signed an informed consent form.

Clinical data extraction

Medical records of all patients were reviewed. Demographic data were considered for analysis, including gender, age, disease course and smoke state; clinical features, laboratory tests, related Antibodies(Antinuclear antibodies, Anti-SCL-70 antibodies, Anti-centromere antibodies and Rheumatoid factor) and pulmonary function tests (diffusing capacity for carbon monoxide and forced vital capacity). Neutrophil-to-leukocyte ratio (NLR) was calculated as the absolute count of neutrophils divided by the absolute count of lymphocytes. Eosinophil-to-leukocyte ratio(ELR) was calculated by dividing the absolute number of eosinophils by the absolute number of lymphocytes. C-reactive protein–to-albumin ratio (CAR) was calculated as C-reactive protein (mg/L) divided by albumin (g/L). Red blood cell distribution width–to-albumin ratio (RAR) was calculated by dividing the red blood cell distribution width (FL) by albumin (g/L). The prognostic nutritional index (PNI) was calculated as albumin (g/L) + 5 × peripheral blood lymphocyte count (×109/L). Studiesshowed that the above indicators are important in the diagnosis and assessment of CTD-ILD and therefore included in our study [15]. Treatment information, including previous or current use of steroids or suppressive drugs. The survival status, survival time of living patients, time of death, and cause of death of deceased patients were followed up by telephone calls.

CT examination

A Lightspeed VCT XT 64-row spiral CT from GE was used, and the CT scan was performed under maximum end-inspiratory breath-hold. The scan was performed in a supine position from the base to the apex of the lungs. Scanning parameters: bulb voltage 120 kV, bulb current 200 mA, layer thickness 1 mm, detector combination 64 × 0.625 mm, pitch 0.984:1, reconstruction of high-resolution lung window (window position – 600, window width 1200), soft tissue window (window position 40, window width 300).

CT assessment

We recorded the presence or absence of the following signs: consolidation, honeycombing, ground glass opacity (GGO), reticular opacities, stretched bronchiectasis, mosaicism, thickening of pulmonary arteries, enlarged lymph nodes, and esophageal dilation. On the basis of the HRCT findings, the distribution was determined: bronchovascular bundle distribution, subpleural distribution, or both. Each patient is evaluated by the radiologist and a predominant HRCT presentation is identified:

1. UIP: honeycomb lung (subpleural/basal predominant), reticular shadows, traction bronchiectasis, no ground glass shadows predominant.
2. NSIP: Ground glass shadows with fine reticular shadows, evenly distributed, no honeycomb lungs, mild bronchiectasis.
3. OP: Patchy solid lesions (bronchovascular bundles/subpleural distribution), may be accompanied by counter halo sign, no cellulitis.
4. LIP: Diffuse ground glass shadows, centrilobular nodules, thin-walled cysts, enlarged lymph nodes, no fibrosis. Representative HRCT images show on the Supplementary Material 2.

Scoring of CT findings

Warrick score of HRCT of the lungs: In the supine position, the CT scan was performed with a layer thickness of 1 mm and a layer spacing of 10 mm; the CT department specialist scored the lungs according to their manifestations (GGO of 1 point, irregular pleural margins of 2 points, pulmonary septal and subpleural lines of 3 points, honeycomb-like changes of 4 points, and subpleural cysts of 5 points). The severity of the disease was assessed by assigning a value to the degree of disease severity, and a score for the extent of disease involvement was determined based on the number of bronchial segments (1–18). (1–3 segments, 1 points; 4–9 segments, 2 points; and > 9 segments, 3 points), and the total of the two scores was added together to form a total score of 0–30 points. These parameters were scored by 2 radiologists. If there was discrepancy, another radiologist who was unaware of the study decided on score.

Statistical analysis

Our data were analyzed using IBM SPSS (version 26.0). Quantitative data were expressed as mean ± standard deviation (normally and near-normally distributed data) or median (interquartile range (IQR); non-normally distributed data), and qualitative data were expressed as numbers and percentages. Given the relatively modest sample size, employing non-parametric methods were used to compare multiple groups. Logistic regression analysis was used to identify risk factors for severe ILD patients, and ROC curves were used to calculate the area under the curve and to analyze the cut-off values for the highest sensitivity and specificity.

The two-step cluster analysis (TCA) was performed to obtain an exploratory cluster analysis using a combination of different types of variables. Other cluster analysis approaches were considered unsuitable as they rely on either continuous or categorical data (hierarchical clustering) or on a preset number of clusters to be distilled (K-means cluster analysis). TCA was developed from BIRCH algorithm and is suitable for large datasets that contain both categorical and/or continuous variables. First, the objects were assigned to "preclusters" and then the preclusters were clustered using hierarchical clustering methods. To check the quality of the clustering, the silhouette measure of cluster cohesion and separation was used. The silhouette measure ranged between $-1 \le 0 \le 1$ and higher values indicated a better clustering structure. More explicitly, values > 0.5 were considered a sign of reasonable structure, and values > 0.7 were regarded as an indicator of strong structure.

We performed variable screening prior to two-step clustering by first including all relevant variables in the initial clustering model, automatically generating a clustered outcome variable, excluding variables with $p > 0.05$ (considered variables that did not contribute to the clustering result) through one-way ANOVA, and 5 variables were eventually included in the cluster analysis model, including Warrick score ≥ 15, anti-SCL-70 antibody, HRCT feature, white blood cell and CAR. Four clusters were obtained for the final analysis. To determine whether there were any significant differences between the clusters, the 1-way ANOVA was used for continuous variables and the chi-square test was used for categorical variables. The probability of survival of the 4 clusters of patients was determined by using the Kaplan-Meier method, and the log-rank test was used to compare curves of probability. The Cox regression model was used to analyze prognostic factors affecting SSc-ILD patients. Statistical significance was defined as $p < 0.05$.

Result

Demographics, manifestations and major organ involvement

In this study, 103 patients with SSc-ILD were enrolled. The disease onset occurred at 54 ± 12.5 years of age, and the median disease duration was 36 months (IQR 12–84 months). The sex ratio in our cohort (male: female) was 1:5.44. Cutaneous involvement was the most prevalent manifestation, occurring in 97.1% of patients ($n = 100$). Furthermore, 80.6% of patients had esophageal involvement ($n = 83$), 27.2% coughed ($n = 28$), 32.0% had chest tightness ($n = 30$), and 35.0% had dyspnea ($n = 36$). The positivity rate for antinuclear antibodies was the highest, reaching 96.1% ($n = 99$). Among patients with SSc-specific antibodies, 48.5% tested positive for anti-SCL-70 antibodies ($n = 50$); 15.5% tested positive for ACA antibodies ($n = 16$). The mean value of Warrick score was 14.4 ± 6.3, 86.4% of patients had lung nodules($n = 89$), 20.4% had lung consolidation($n = 21$), 88.3% had traction bronchiectasis($n = 91$), 89.3% had reticular pattern($n = 92$), 7.8% had mosaic attenuation ($n = 8$), 37.9% had thickened pulmonary arteries($n = 39$), 43.7% had enlarged lymph nodes($n = 45$), and 49.5% had esophageal dilation ($n = 51$). Regardless of the HRCT feature, subpleural lesions dominated the lesion distribution. In our study, the NSIP ($n = 65$,63.1%) was the most common HRCT feature in SSc-ILD patients. NSIP patients had a reticular pattern in 98.5% of cases, significantly higher than the other groups ($p = 0.000$). Patients with UIP had the highest percentage of cough symptoms, the highest percentage of SCL-70 positivity, and the highest Warrick score. The best prognosis was seen among patients with NSIP, while the worst was seen among those with UIP. Specific data are shown in Table 1.

Clinical prediction models

A warrick score ≥ 15 was used as a criterion for classifying ILDs as severe. A comparison showed that patients with severe disease had a longer disease duration ($p = 0.002$), a higher proportion of males ($p = 0.022$), a higher proportion of smokers ($p = 0.034$), and cough symptoms ($p = 0.030$) were more prevalent than those with non-severe disease. NSIP was the most prevalent HRCT feature in both groups, but UIP was significantly more prevalent in the severe group ($p = 0.000$) than in the non-severe group. Compared to the non-severe group, the severe group had a significantly higher platelet count ($p = 0.017$). The severe group had a poor prognosis ($p = 0.021$) and a high mortality rate. Specific data are shown in Table 2. To understand the risk factors leading to severe ILD, we constructed a clinical prediction model through logistic analysis. Variable screening was first performed, and variables with $p < 0.05$ were screened by one-way logistic regression, and these were included

Table 1 Characteristics of the SSc-ILD patients in the different HRCT feature

Variables	All patients($n=103$)	NSIP($n=65$)	LIP($n=18$)	OP($n=10$)	UIP($n=10$)	*p*
Demographic characteristics						
Gender(M/F)	16/87	11/54	3/15	2/8	0/10	0.304
Age(years)	54 ± 12.5	53 ± 12.8	55 ± 13.1	56 ± 12.3	52 ± 10.8	0.951
Disease duration(months)	36(12,84)	36(12,84)	24(12,96)	54(20,141)	78(36,126)	0.223
Smoke(n,%)	9(8.7%)	6(9.2%)	2(11.1%)	1(10%)	0(0%)	0.575
Clinical manifestations						
Cutaneous involvement (n,%)	100(97.1%)	100(100.0%)	16(88.9%)	10(100.0%)	9(90.0%)	0.323
Esophageal involvement (n,%)	83(80.6%)	54(83.1%)	15(83.3%)	7(70.0%)	7(70.0%)	0.544
Cough(n,%)	28(27.2%)	16(24.6%)	2(11.1%)	4(40.0%)	6(60.0%)	0.037*
Chest tightness(n,%)	33(32.0%)	20(30.8%)	2(11.1%)	6(60.0%)	5(50.0%)	0.028
Dyspnea(n,%)	36(35.0%)	24(36.9%)	2(11.1%)	5(50.0%)	5(50.0%)	0.060
Raynaud's symptoms(n,%)	93(90.3%)	61(93.8%)	16(88.9%)	7(70.0%)	9(90.0%)	0.129
Finger ulcers(n,%)	12(11.7%)	6(9.2%)	3(16.7%)	1(10.0%)	2(20.0%)	0.679
Laboratory examinations						
FVC actual/forecast(%)	75.69 ± 18.44	76.90 ± 18.07	81.06 ± 21.98	63.83 ± 17.91	67.69 ± 15.67	0.324
DLCO actual/forecast(%)	56.11 ± 18.07	54.77 ± 15.83	62.06 ± 23.28	78.83 ± 27.30	47.00 ± 12.84	0.125
ANA positive(n,%)	99(96.1%)	62(95.4%)	17(94.4%)	10(100%)	10(100%)	0.647
SCL-70 positive(n,%)	50(48.5%)	35(53.8%)	6(33.3%)	1(10%)	8(80%)	0.003*
ACA positive(n,%)	16(15.5%)	7(10.8%)	6(33.3%)	1(10%)	2(20%)	0.169
IgG(g/L)	15.45(11.83,18.80)	15.60(13.00,18.10)	12.90(10.35,15.78)	15.60(12.25,20.50)	19.10(10.68,22.20)	0.177
IgA(g/L)	2.76(2.05,4.06)	2.71(2.05,4.05)	2.74(1.71,3.09)	2.78(1.89,4.70)	2.79(2.29,5.13)	0.597
IgM(g/L)	1.39(0.92,1.76)	1.47(0.95,1.95)	1.12(0.70,1.42)	1.55(0.84,2.07)	1.18(0.82,1.53)	0.068
IgE(IU/ml)	40.80(19.40,122.50)	39.80(19.30,196.75)	30.35(20.00,50.95)	50.50(41.30,101.25)	41.45(19.33,181.25)	0.592
NLR	2.60(1.89,3.84)	2.38(1.84,3.63)	2.61(1.88,3.81)	3.59(2.21,5.95)	3.65(1.88,6.24)	0.181
ELR	0.04(0.01,0.10)	0.04(0.01,0.09)	0.08(0.03,0.14)	0.05(0.01,0.12)	0.03(0.01,0.06)	0.318
CAR	0.08(0.02,0.29)	0.06(0.02,0.30)	0.08(0.03,0.11)	0.10(0.04,0.59)	0.15(0.03,0.43)	0.573
RAR	0.38(0.34,0.46)	0.39(0.34,0.45)	0.36(0.32,0.48)	0.41(0.35,0.56)	0.38(0.34,0.56)	0.460
PNI	44.65(40.35,48.90)	44.50(40.70,48.35)	46.33(42.35,49.13)	43.65(33.71,48.16)	46.40(34.75,49.20)	0.609
HRCT						
Warrick score	14.4 ± 6.3	13.0 ± 5.1	12.8 ± 5.6	15.7 ± 5.1	24.5 ± 6.4	0.000***
Lung nodule(n,%)	89(86.4%)	56(86.2%)	15(83.3%)	9(90%)	9(90%)	0.945
Lung consolidation(n,%)	21(20.4%)	8(12.3%)	2(11.1%)	10(100%)	1(10%)	0.000***
Traction bronchiectasis(n,%)	91(88.3%)	58(89.2%)	13(72.2%)	5(27.8%)	10(100%)	0.038*
Reticular pattern(n,%)	92(89.3%)	64(98.5%)	11(61.1%)	8(80%)	9(90%)	0.000***
Mosaic attenuation(n,%)	8(7.8%)	5(7.7%)	2(11.1%)	0(0%)	1(10%)	0.587
Thickening of the pulmonary arteries(n,%)	39(37.9%)	27(41.5%)	4(22.2%)	6(60%)	2(20%)	0.117
Esophageal dilatation(n,%)	51(49.5%)	35(53.8%)	5(27.8%)	6(60%)	5(50%)	0.216
Enlarged lymph nodes(n,%)	45(43.7%)	33(50.8%)	4(22.2%)	4(40%)	4(40%)	0.167
Range(n,%)						
Broncho vascular bundle	2(1.9%)	0(0%)	2(11.1%)	0(0%)	0(0%)	0.188
Subpleural areas	98(95.1%)	63(96.9%)	16(88.9%)	9(90%)	10(100%)	-
Both	3(2.9%)	2(3.1%)	0(0%)	1(10%)	0(0%)	-
Pleuritis(n,%)	18(17.5%)	9(13.8%)	2(11.1%)	5(50.0%)	2(20.0%)	0.081
Treatment						
Past or current Steroids (n,%)	100(97.1%)	63(96.9%)	17(94.4%)	10(100.0%)	10(100.0%)	0.673
Past or current Cyclophosphamide(n,%)	3(2.9%)	2(3.1%)	0(0.0%)	0(0.0%)	1(10.0%)	0.429
Past or current Mycophenolate Mofetil (n,%)	60(58.3%)	41(63.1%)	10(55.6%)	5(50.0%)	4(40.0%)	0.512
Past or current Pirfenidone (n,%)	9(8.7%)	8(12.3%)	0(0.0%)	0(0.0%)	1(10.0%)	0.108
Past or current Nintedanib (n,%)	9(8.7%)	6(9.2%)	1(5.6%)	2(20.0%)	0(0.0%)	0.346

Table 1 (continued)

Variables	All patients($n=103$)	NSIP($n=65$)	LIP($n=18$)	OP($n=10$)	UIP($n=10$)	*p*
Prognosis						
Survivor(n,%)	84(81.6%)	59(90.8%)	13(72.2%)	6(60.0%)	6(60.0%)	0.016*

Esophageal involvement includes Gastroesophageal Reflux Disease (GERD)/Reflux Esophagitis (RE), FVC: forced vital capacity, DLCO: diffusing capacity for carbon monoxide, ANA: Anti-nucleosome antibodies, SCL-70:anti-SCL-70 antibody, ACA: anti-centromere antibody, RF: rheumatoid factor, IgA/M/E/G: immunoglobulin A/M/E/G, ELR: eosinophil to lymphocyte ratio, CAR: c-reactive protein to albumin ratio, RAR: red cell distribution width to albumin ratio, PNI: prognostic nutritional index, UIP: Usual interstitial pneumonia, NSIP: Nonspecific interstitial pneumonia, OP: Organizing Pneumonia, LIP: lymphocytic interstitial pneumonia, Characteristics were compared among the group using the one-way ANOVA or Mann-Whitney *U* as appropriate. *$P<0.05$ ***$P<0.001$

Table 2 Characteristics of the SSc-ILD patients in the different Warrick score

Variables	Warrick < 15($n=78$)	Warrick ≥ 15($n=25$)	*p*
Disease duration(months)	33(12,72)	96(36,174)	0.002*
Gender(M/F)	8/70	8/17	0.022*
Smoke(n,%)	4(5.1%)	5(20.0%)	0.034*
Cough(n,%)	17(21,8%)	11(44.0%)	0.030*
Chest tightness(n,%)	22(28.2%)	11(44.0%)	0.141
Dyspnea(n,%)	25(32.1%)	11(44.0%)	0.276
Basophil percentage(%)	0.32 ± 0.28	0.17 ± 0.14	0.035*
Lung nodule(n,%)	68(87.2%)	21(84.0%)	0.945
Lung consolidation(n,%)	17(21.8%)	4(16.0%)	0.531
Traction bronchiectasis(n,%)	67(85.9%)	24(96.0%)	0.312
Reticular pattern(n,%)	70(89.7%)	22(88.0%)	0.806
Mosaic attenuation(n,%)	8(10.3%)	0(0.0%)	0.216
Thickening of the pulmonary arteries(n,%)	32(41.0%)	7(28.0%)	0.243
Esophageal dilatation(n,%)	37(47.4%)	14(56.0%)	0.456
Enlarged lymph nodes(n,%)	37(47.4%)	8(32.0%)	0.176
Range(n,%)			
Broncho vascular bundle	2(2.6%)	0(0.0%)	0.538
Subpleural areas	74(94.9%)	24(96.0%)	-
Both	2(2.6%)	1(4.0%)	-
Pleuritis(n,%)	12(15.4%)	6(24.0%)	0.324
Imaging type(n,%)			
UIP	2(2.6%)	8(32.0%)	0.000***
NSIP	56(71.8%)	9(36.0%)	-
OP	7(9.0%)	3(12.0%)	-
LIP	13(16.7%)	5(20.0%)	-
FVC actual/forecast(%)	79.70 ± 17.72	66.01 ± 16.91	0.655
DLCO actual/forecast(%)	59.76 ± 15.05	47.52 ± 21.88	0.051
SCL-70 positive(n,%)	37(47.4%)	13(52.0%)	0.691
PLT(10^9/L)	262(199,320)	278(194,363)	0.017*
Survivor(n,%)	68(87.2%)	16(64.0%)	0.021*

UIP: Usual interstitial pneumonia, NSIP: Nonspecific interstitial pneumonia, OP: Organizing Pneumonia, LIP: lymphocytic interstitial pneumonia, FVC: forced vital capacity, DLCO: diffusing capacity for carbon monoxide, SCL-70:anti-SCL-70 antibody, PLT: platelet, Characteristics were compared among the group using the one-way ANOVA or Mann-Whitney *U* as appropriate. *$P<0.05$ ***$P<0.001$

in a multifactorial logistic regression model. A multifactor logistic regression model was constructed with the following variables: smoking, gender, disease duration, joint pain, erythrocytes and basophil percentage were included in a multivariate logistic regression model, and it was found that only disease duration was a predictor of risk for a Warrick score of ≥ 15 in patients with ILD, and that with each month of increased disease duration, the Warrick score ≥ 15 risk increased 1.016-fold. Erythrocyte, basophil percentage, and sex were protective factors for ILD patients; for every 1% increase in erythrocyte count, there was a 0.206-fold reduction in the risk of a Warrick score ≥ 15 in patients with ILD and for every 1% increase in basophil percentage, there was a 0.021-fold increase in risk of Warrick score ≥ 15 in ILD patients. The risk of a Warrick score ≥ 15 in male patients with SSC-ILD was 1.122-fold higher than in female patients.

The above variables were subjected to ROC analysis, excluding those with $p>0.05$, and statistically analysed for the best critical values of the remaining variables

with the highest sensitivity and specificity. Patients had a high likelihood of having a Warrick score greater than 15 when they had a disease duration of > 102 months and a basophil percentage of < 0.15%, the patient had a high likelihood of having severe ILD. Result show on the Supplementary Material 3.

For a more graphic representation of the above conclusion, we use the matrix model. Sex male or female, disease duration ≤ 102months or > 102months, and Basophil percentage ≥ 0.15% or < 0.15% were combined to generate a matrix, and then, the relative risk probability of severe ILD in each grid was calculated. The 95% confidence intervals of the probabilities were shown in the parentheses. The red grid indicates high risk, with a > 50% incidence of severe ILDs. The yellow grid indicates moderate risk, with a prevalence of severe ILDs between 49% and 10%. The green grid indicates low risk, with a < 10% incidence of severe ILDs. The matrix prediction model is shown in Fig. 1.

Cluster analysis

The clustering of individuals based on the 5 selected variables (including Warrick score ≥ 15, anti-SCL-70 antibody, HRCT feature, white blood cell and CAR) yielded an optimal number of four clusters. The average contour value of clustering was 0.6, indicating good clustering quality. The characteristics of clusters 1–4 are summarized in Table 3**and** Fig. 2.

Cluster 1 ($n = 23$, 24.5%), had a 9.5% male patient ratio, a mean disease duration of 36 months, and 8.7% smokers. All patients had cutaneous involvement as the most common clinical sign. As compared to the other three clustered groups, FVC ($p = 0.039$), DLCO ($p = 0.048$) were significantly higher. The mean value of Krebs von den Lungen-6 antigen (KL-6) was 450 U/ml. There were 34.8% of patients with SCL-70 antibodies and 26.1% with ACA antibodies. In HRCT, the mean value of Warrick score was 11.57 ± 3.59 and all less than 15;the most common signs were lung nodule and traction bronchiectasis (82.6%), followed by reticular pattern (69.6%); the least common sign was mosaic attenuation, which was seen in only 13.0% of patients. The proportion of lung consolidation was significantly higher than in the other subgroups ($p = 0.040$); and LIP was the predominant HRCT feature in this group, with a lower percentage of UIP pattern, 8.7% (2/23). 99% of patients were treated with Mycophenolate Mofetil (MMF), while only 4.3% of patients were treated with Nintedanib. Of the patients treated in the four subgroups, 69.6% survived.

Cluster 2 ($n = 23$, 24.5%), which had a higher proportion of male patients ($p = 0.040$) and disease duration ($p = 0.029$) than the other clusters, the disease duration are 60(36,180) months, had lower FVC and DLCO values. The mean KL-6 value was 909 U/ml. 52.2% of patients had anti-SCL-70 antibodies positive, and 8.7% had anti-ACA antibodies positive. The mean Warrick score was as high as 23.43 ± 4.34 and all ≥ 15, reflecting the severity of lung disease in this group, which was an important reason for the poorer prognosis and highest mortality rate. Traction bronchiectasis was the most common sign (95.7%); no mosaic attenuation was observed. A dominant feature of HRCT was NSIP and the proportion of UIP patterns amounted to 26.1% (6/23), a higher proportion than in the other clustering groups. Cluster 2 had the highest percentage of patients using MMF (78.3%, $p = 0.021$). Patients in this group had the worst prognosis ($p = 0.007$) and the highest mortality rate.

Cluster 3 ($n = 20$, 21.3%), The median disease duration was 36 months (24, 69), which was similar to the overall median. All patients were negative for SCL-70 antibodies,

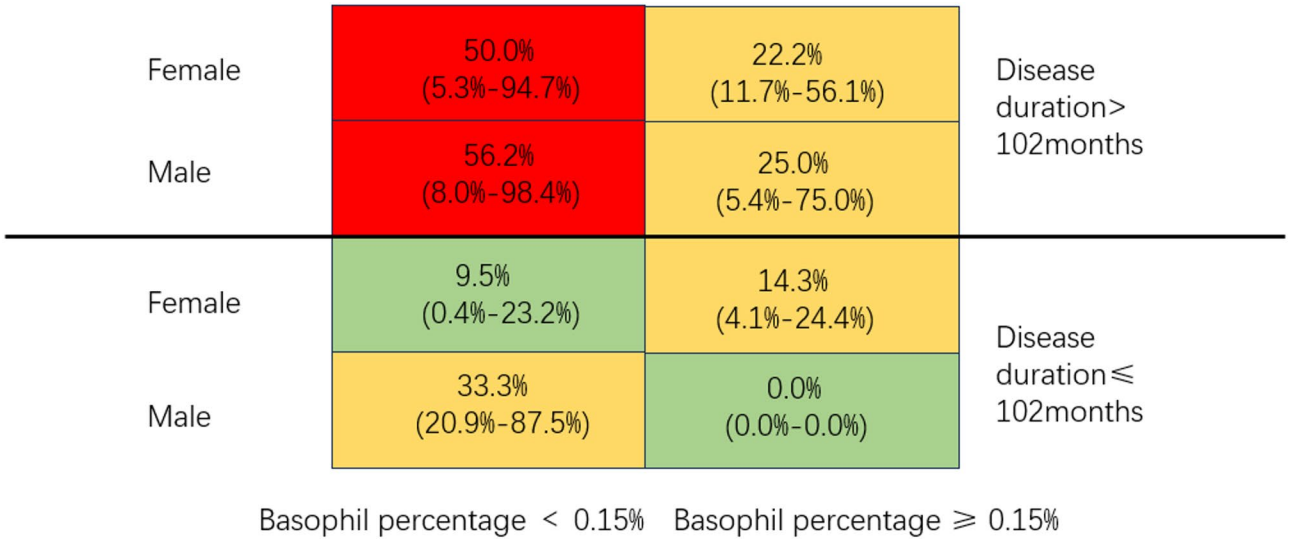

Fig. 1 Matrix prediction model

Table 3 Characteristics of the SSc-ILD patients in the different cluster group

Variables	All patients($n=103$)	Cluster 1($n=23$)	Cluster 2($n=23$)	Cluster 3($n=20$)	Cluster 4($n=28$)	*p*
Demographic characteristics						
Gender(M/F)	16/87	2/21	8/15	1/19	4/24	0.040*
Age(years)	54 ± 12.5	55 ± 12.8	53 ± 13.3	51 ± 10.2	54 ± 13.3	0.753
Disease duration(months)	36(12,84)	36(12,108)	60(36,180)	36(24,69)	27(12,57)	0.029*
Smoke(n.%)	9(8.7%)	2(8.7%)	5(21.7%)	0(0.0%)	2(7.1%)	0.064
Clinical manifestations						
Cutaneous involvement (n,%)	100(97.1%)	23(100.0%)	22(95.7%)	19(95.0%)	27(96.4%)	0.569
Esophageal involvement (n,%)	83(80.6%)	22(96.7%)	21(91.3%)	18(90.0%)	22(78.6%)	0.121
Cough(n,%)	28(27.2%)	5(21.7%)	10(43.5%)	4(20.0%)	8(28.6%)	0.291
Chest tightness(n,%)	33(32.0%)	7(30.4%)	9(39.1%)	6(30.0%)	9(32.1%)	0.910
Dyspnea(n,%)	36(35.0%)	6(26.1%)	9(39.1%)	8(40.0%)	11(39.3%)	0.707
Raynaud's symptoms(n,%)	93(90.3%)	19(82.6%)	21(91.3%)	19(95.0%)	27(96.4%)	0.318
Finger ulcers(n,%)	12(11.7%)	2(8.7%)	3(13.0%)	4(20.0%)	2(7.1%)	0.544
Laboratory examinations						
FVC actual/forecast(%)	75.69 ± 18.44	83.20 ± 21.21	64.06 ± 15.37	76.15 ± 17.70	81.64 ± 17.18	0.039*
DLCO actual/forecast(%)	56.11 ± 18.07	68.86 ± 19.85	47.70 ± 22.58	53.46 ± 12.33	59.92 ± 14.03	0.048*
ANA positive(n,%)	99(96.1%)	23(100.0%)	20(87.0%)	19(95.0%)	28(100.0%)	0.062
SCL-70 positive(n,%)	50(48.5%)	8(34.8%)	12(52.2%)	0(0.0%)	28(100.0%)	0.000***
ACA positive(n,%)	16(15.5%)	6(26.1%)	2(8.7%)	5(25.0%)	2(7.1%)	0.173
RF positive(n,%)	26(25.2%)	2(8.7%)	6(26.1%)	10(50.0%)	5(17.9%)	0.015*
IgG(g/L)	15.45(11.83,18.80)	13.90(10.35,19.48)	15.30(11.90,18.80)	16.35(11.65,22.33)	15.40(13.00,17.20)	0.696
IgA(g/L)	2.76(2.05,4.06)	2.75(1.88,4.43)	2.86(2.33,4.53)	2.81(2.24,3.98)	2.60(2.06,4.08)	0.840
IgM(g/L)	1.39(0.92,1.76)	1.33(0.84,1.67)	1.25(0.87,1.70)	1.60(0.98,2.34)	1.35(0.94,1.73)	0.236
IgE(IU/ml)	40.80(19.40,122.50)	43.85(26.55,71.53)	47.00(19.40,99.80)	37.35(20.08,161.13)	35.90(18.50,264.50)	0.966
NLR	2.60(1.89,3.84)	2.75(2.19,3.82)	3.31(1.64,4.48)	2.52(1.67,4.30)	2.56(1.84,3.50)	0.722
ELR	0.04(0.01,0.10)	0.07(0.03,0.13)	0.03(0.01,0.06)	0.02(0.00,0.10)	0.04(0.02,0.10)	0.216
CAR	0.08(0.02,0.29)	0.08(0.05,0.11)	0.24(0.01,0.64)	0.04(0.01,0.09)	0.07(0.02,0.35)	0.133
RAR	0.38(0.34,0.46)	0.36(0.33,0.53)	0.39(0.35,0.49)	0.39(0.35,0.47)	0.35(0.33,0.41)	0.175
PNI	44.65(40.35,48.90)	43.00(38.50,48.90)	45.25(41.70,49.20)	44.33(42.85,46.10)	46.83(40.51,52.95)	0.303
IL-6(pg/ml)	13.79 ± 21.41	13.24 ± 11.74	6.50 ± 6.23	5.61 ± 3.23	14.91 ± 16.48	0.691
KL-6(U/ml)	1285(1003,1895)	450(450,450)	1533(1182,2387)	-	909(721,721)	0.061
FER(μg/L)	62.30(16.40,96.05)	5.40(5.40,5.40)	95.86(57.70,95.86)	30.40(13.60,30.40)	96.23(19.20,96.23)	0.213
HRCT						
Warrick score	14.4 ± 6.3	11.57 ± 3.59	23.43 ± 4.34	11.52 ± 3.16	10.90 ± 2.52	0.000***
Warrick ≥ 15(n,%)	23(22.3%)	0(0.0%)	23(100%)	0(0.0%)	0(0.0%)	0.000***
Lung nodule(n,%)	89(86.4%)	19(82.6%)	19(82.6%)	18(90.0%)	25(89.3%)	0.808
Lung consolidation(n,%)	21(20.4%)	10(43.5%)	4(17.4%)	2(10.0%)	5(17.9%)	0.040*
Traction bronchiectasis(n,%)	91(88.3%)	19(82.6%)	22(95.7%)	17(85.0%)	26(92.9%)	0.405
Reticular pattern(n,%)	92(89.3%)	16(69.6%)	20(87.0%)	19(95.0%)	28(100.0%)	0.003*
Mosaic attenuation(n,%)	8(7.8%)	3(13.0%)	0(0.0%)	2(10.0%)	3(10.7%)	0.184
Thickening of the pulmonary arteries(n,%)	39(37.9%)	9(39.1%)	6(26.1%)	10(50.0%)	8(28.6%)	0.331
Esophageal dilatation(n,%)	51(49.5%)	10(43.5%)	12(52.2%)	12(60.0%)	12(42.9%)	0.005*
Enlarged lymph nodes(n,%)	45(43.7%)	9(39.1%)	8(34.8%)	7(35.0%)	15(53.6%)	0.474
Range(n,%)						
Broncho vascular bundle	2(1.9%)	2(8.7%)	0(0.0%)	0(0.0%)	0(0.0%)	0.238
Subpleural areas	98(95.1%)	20(87.0%)	22(95.7%)	19(95.0%)	28(100.0%)	-
Both	3(2.9%)	1(4.3%)	1(4.3%)	1(5.0%)	0(0.0%)	-
Imaging typing(n,%)						
UIP	10(9.7%)	2(8.7%)	6(26.1%)	0(0.0%)	0(0.0%)	0.000***
NSIP	65(63.1%)	1(4.3%)	9(39.1%)	20(100.0%)	28(100.0%)	-
OP	10(9.7%)	7(30.4%)	3(13.0%)	0(0.0%)	0(0.0%)	-
LIP	18(17.5%)	13(56.5%)	5(21.7%)	0(0.0%)	0(0.0%)	-

Table 3 (continued)

Variables	All patients($n=103$)	Cluster 1($n=23$)	Cluster 2($n=23$)	Cluster 3($n=20$)	Cluster 4($n=28$)	*p*
Pleuritis(n,%)	18(17.5%)	6(26.1%)	6(26.1%)	3(15.0%)	2(7.1%)	0.192
Treatment						
Past or current Steroids(n,%)	100(97.1%)	22(95.7%)	23(100.0%)	20(100.0%)	27(96.4%)	0.475
Past or current Cyclophosphamide(n,%)	3(2.9%)	0(0.0%)	1(4.3%)	1(5.0%)	1(3.6%)	0.620
Past or current Mycophenolate Mofetil (n,%)	60(58.3%)	9(39.1%)	18(78.3%)	10(50.0%)	19(67.9%)	0.021*
Past or current Pirfenidone (n,%)	9(8.7%)	0(0.0%)	4(17.4%)	1(5.0%)	4(14.3%)	0.053
Past or current Nintedanib (n,%)	9(8.7%)	1(4.3%)	3(13.0%)	0(0.0%)	5(17.9%)	0.071
Prognosis						
Survivor(n,%)	84(81.6%)	16(69.6%)	15(65.2%)	18(90.0%)	27(96.4%)	0.007*

Esophageal involvement includes Gastroesophageal Reflux Disease (GERD)/Reflux Esophagitis (RE); FVC: forced vital capacity, DLCO: diffusing capacity for carbon monoxide, ANA: Anti-nucleosome antibodies, SCL-70:anti-SCL-70 antibody, ACA: anti-centromere antibody, RF: rheumatoid factor, IgA/M/E/G: immunoglobulin A/M/E/G, ELR: eosinophil to lymphocyte ratio, CAR: c-reactive protein to albumin ratio, RAR: red cell distribution width to albumin ratio, PNI: prognostic nutritional index, IL-6: interleukin-6, KL-6: Krebs von den Lungen-6 antigen, FER: ferritin, UIP: Usual interstitial pneumonia, NSIP: Nonspecific interstitial pneumonia, OP: Organizing Pneumonia, LIP: lymphocytic interstitial pneumonia, Characteristics were compared among the group using the one-way ANOVA or Mann-Whitney *U* as appropriate. *$P<0.05$ ***$P<0.001$

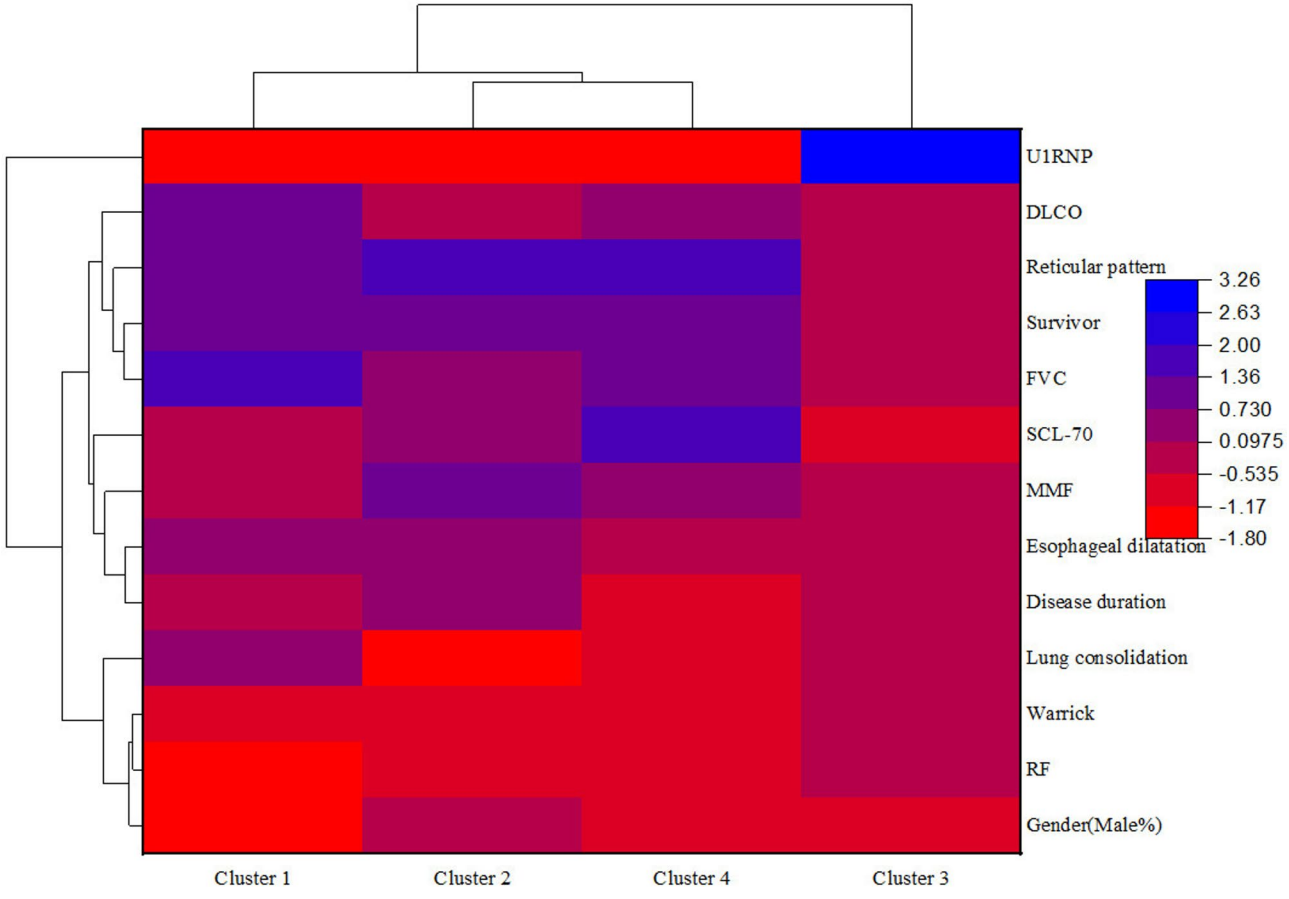

Fig. 2 Heatmap showing the clinical characteristics in each cluster

the percentage of RF antibody positivity was higher than the other groups ($p=0.015$), 25.0% of patients were positive for ACA antibodies, and the mean Warrick score was 11.52 ± 3.16, which was at a lower level; the percentage of patients with esophageal dilatation sign was higher than the other groups ($p=0.005$); all patients HRCT features were NSIP and no patients presented with UIP; 50% of patients were on MMF ($n=10$) and 5% were on

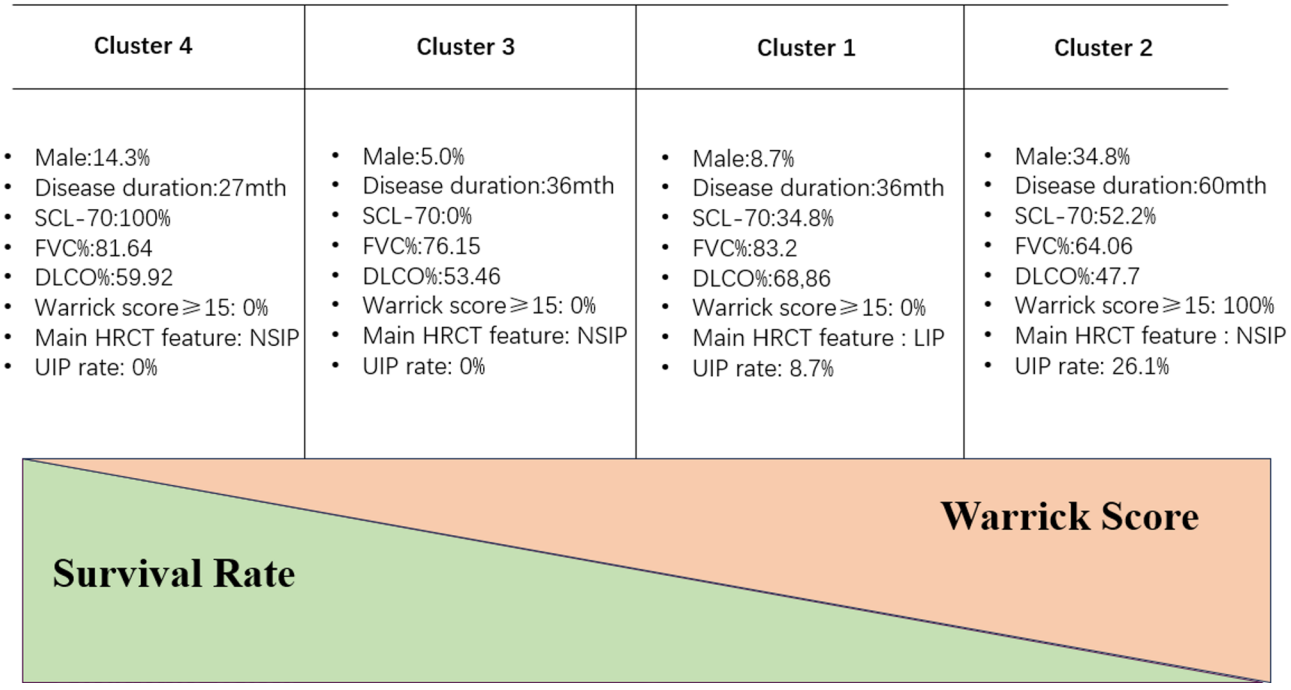

Fig. 3 Clinical characteristics in each cluster

Pirfenidone ($n=1$). There was a good prognosis for 90% of patients.

Cluster 4 ($n=28$, 29.8%), The duration of the disease was shorter, with a median of 27 months (12, 57), significantly shorter than in the other groups ($p=0.029$), all patients were positive for SCL-70 antibodies, and the mean Warrick score was 10.90 ± 2.52, which was lower than the other group ($p=0.000$). There was a reticular pattern change in all patients; all patients' HRCT features were NSIP and no patients presented with UIP; 67.9% of patients were on MMF, 14.3% on Pirfenidone ($n=4$); 17.9% on Nintedanib ($n=5$). This group of patients had the best prognosis ($p=0.007$).

Figure 3 shows the key clinical features of Clusters 1–4 are presented graphically, including the proportion of males, disease duration, SCL −70 antibody status, FVC and DLCO values, Warrick score ≥ 15, key HRCT features, and UIP proportion information. The key clinical features of each cluster are displayed centrally to facilitate visual comparison of differences between clusters. Through comparison, it can be found that Cluster 2 has a relatively higher proportion of males, longer disease duration, 100% Warrick score ≥ 15, and a higher proportion of UIP, which indicate that patients in this cluster have a more severe disease and may have a poorer prognosis; whereas, patients in Cluster 4 have a shorter duration of the disease, are positive for SCL-70 antibody, and have NSIP as the main HRCT feature, which suggests that patients in this cluster have a more severe disease and may have a poorer prognosis. The HRCT characteristics of Cluster 4 patients were NSIP, suggesting that patients in this cluster may have a relatively good prognosis.

Survival analyses

During the follow-up period, 18.4% of patients (19/103) died. Mortality in cluster 2 was 34.8%, significantly higher than that in clusters 1 (30.4%, Pc < 0.000), 3(10.0%, Pc = 0.026) and 4(3.6%, Pc = 0.011). In Kaplan-Meier analysis, cluster 2 patients had a greater risk of death than patients in the other three clusters (log-rank $P=0.012$). Kaplan-Meier curves are shown in Supplementary Material 4.

Variables with $p<0.1$ from one-way cox regression analysis were included in a multifactor COX regression model, the multifactor Cox regression model in SSC-ILD was constructed with the following factors: age, albumin, C-reactive protein, lymphocyte percentage, absolute neutrophil value, immunoglobulin M(IgM), RAR and PNI. It was ultimately found that albumin was a protective factor for SSC-ILD patients (HR:0.219, 95%CI:0.062–0.770, $p=0.018$), whereas age (HR:1.169, 95%CI:1.010–1.352, $p=0.036$) and IgM(HR:1.077, 95%CI:1.007–1.893, $p=0.040$) were risk factors. The results of the COX analysis for the other clustered groups are shown in Supplementary Material 5.

Discussion

SSC-ILD can occur before clinical symptoms appear, and HRCT, the gold standard for diagnosing ILD, has not been well integrated with clinical and laboratory data. Previous work stratifying SSc-ILD has focused on multivariate predictive models integrating clinical, serological and imaging data. Kuwana et al. [16] developed a prognostic model combining lung function decline, HRCT range, and autoantibody profile to predict progression of ILD. Their approach emphasises longitudinal risk prediction and ignores disease heterogeneity. Thus, in this study, we developed a clinical prediction model that combined clinical manifestations, laboratory tests, antibody tests, and HRCT signs. And used cluster analysis to identify four homogeneous groups among 103 SSC-ILD patients with significant differences in clinical manifestations, autoantibody patterns, HRCT signs, and prognosis, in order to guide early detection of ILD and precise treatment.

Previous studies have shown that HRCT supports the classification of histological patterns of ILD [17], and our findings are identical with these established correlations. Our analysis of HRCT indicated that SSc patients mostly suffered from NSIP (63.1%); these findings are consistent with a previous study [18]. NSIP are characterized by confluent GGO and a fine reticular pattern, often posterior and subpleural, usually associated with traction bronchiectasis and bronchiolectasis. Honeycombing, when present, is usually mild [19]. Immunosuppression with corticosteroids ± cytotoxic agents are treatments of choice [20], although there are no clear guidelines regarding glucocorticoids dosage or duration of therapy. In a study of a mixed idiopathic/secondary NSIP cohort, Lee et al. reported that relapses were associated with low initial prednisolone therapy (0.5 mg/kg) as well as shorter duration of therapy (4.7 compared with 7.7 months) prior to steroid cessation [21]. NSIP has a relatively good prognosis [22]. The study found that the lung fibrosis associated with scleroderma is associated with a much better prognosis than that found in idiopathic lung fibrosis, most likely due, in part, to the predominant NSIP histology [23]. By cluster analysis, we found that UIP patients with Warrick score ≥ 15 and positive SCL-70 antibodies had the worst prognosis, and early combination of immunosuppressive as well as antifibrotic medications was recommended.

In this study, we chose the Warrick score to quantify as accurately as possible the degree of physiologic or radiographic abnormality at the time of CT examination. We found a significant association between Warrick score and patient prognosis which consistent with previous studies [24]. Our study found that patients with UIP had the highest Warrick score and the worst prognosis ($p < 0.001$). This is consistent with the results of a previous study of Sjögren's Syndrome combined with ILD [25]. In a study on Rheumatoid Arthritis(RA)it was reported that the UIP pattern was seen in 24% of patients with RA-ILD, and those patients showed worse survival, with a similar disease trajectory to patients with IPF [26]. Since then, multiple studies have found an association between the UIP pattern on HRCT and higher mortality. The above findings suggest that the Warrick score may be useful in predicting imaging patterns and prognosis in patients with ILD. The Warrick score does more than that. Obstructive sleep apnea(OSA) has been reported in 66% of patients with scleroderma with pulmonary involvement [27]. Two different mechanisms have been described. In the presence of ILD, decreased lung volume results in impaired upper respiratory airway stability, as well as increased pharyngeal collapsibility, by reducing the downward expanding force (caudal traction) on the pharynx [28]. A further mechanism for the association of sleep-disordered breathing in patients with ILD, is obesity, which develops in relation to the corticosteroids used for the treatment of ILD [29]. In Sarac's study, a significant relationship was found between the Warrick score and the OSA severity parameters, as determined by polysomnography [30].

In our study, IgM was one of the risk factors for predicting death in patients with SSc-ILD. The relationship between IgM and ILD has also been reported in previous studies. In their study of serum markers of interstitial lung disease associated with primary Sjögren's syndrome, Liu et al. found that higher levels of IgM were associated with the development of ILD [31]. Wang et al. found that IgM in the dermatomyositis (DM) associated ILD (ILD-DM) were higher than that of the DM non-complicating ILD (Non-ILD-DM) ($p = 0.029$) [32]. It was verified that IgM plays a role in the development of ILD, but the exact mechanism of action is not clear.

We demonstrate that basophils are a prognostic marker for survival for SSc-ILD. A previous study found that a detectable number of basophils at baseline was a novel risk factor for early death and the occurrence of acute exacerbation-ILD by analyzing the bronchoalveolar lavage of ILD patients [33]. Cohen has demonstrated in vitro and in mouse experiments that basophils contribute significantly to the development and function of alveolar macrophages in controlling lung homeostasis. In addition, he believes that in pathologies that basophil depletion may be beneficial in attenuating the detrimental immune response by suppressing tumor associated and pro-fibrotic macrophage activity [34]. Our results suggest that this cell type and its role as a prognostic marker should be clarified in the future.

Traditional stratification of SSc relies heavily on skin involvement and autoantibody profiles. The EUSTAR cohort study by Elhai et al. [35] demonstrated that

autoantibody status more strongly predicts organ-specific outcomes than skin subtype. Our findings align with this, as anti-Scl-70-positive clusters exhibited distinct HRCT patterns and poorer prognosis. In our study, we found that patients with UIP had the highest percentage of SCL-70 antibody positivity and the worst prognosis. Previous studies have demonstrated that the risk of major organ involvement in Scl-70 patients, including a higher risk of ILD, renal crisis and myocardial involvement [36]. And Scl-70 positivity has been associated with diffuse skin involvement and a poorer prognosis. We therefore propose the conclusion that SSc patients with anti-SCL-70 antibodies may have a more frequent pattern of UIP and a worse prognosis. Bruni, C. et al. developed the ILD - RISC scoring model [37]. This model incorporates variables such as FVC%, DLCO%, history of digital ulcers, age, and scleroderma autoantibody status. It can effectively predict the presence of systemic sclerosis - associated interstitial lung disease. The model has shown excellent performance in both the derivation cohort and the validation cohort, guiding the decision on whether HRCT is required during follow - up. Future studies should explore integrating the ILD-RISC risk score with our cluster-based subgroups to develop dynamic prediction models which could improve early identification of progressive ILD and guide personalized therapeutic interventions.

LIP is often associated with aberrant B-cell activation, lymphoproliferative lesions, and is most frequently observed in patients with combined sicca syndrome (SSc-sicca syndrome) or other immune disorders [38]. The higher prevalence of LIP in our study may be related to the fact that these patients have a tendency to develop SSc-sicca syndrome but have not yet met the diagnostic criteria. In addition, positivity for anti-centromere antibodies(ACA) is associated with a high ILP ratio, and binding of ACA to antigens on vascular endothelial cells damages the pulmonary microvasculature, causing localized hypoxia, attracting lymphocyte aggregates, triggering inflammation, and contributing to the formation of LIP. The solid features of OP reflect intra-alveolar fibroblast proliferation and inflammatory exudation, which are commonly seen during acute exacerbations of SSc-ILD or during the active inflammatory phase [39]. The patients included in our study were predominantly in the early stages of ILD (median disease duration 36 months). As an inflammatory phenotype, LIP/OP may be more common in the early stages of the disease, whereas the proportion of patients with advanced fibrosis (usual interstitial pneumonia) was relatively lower. Additionally, the high prevalence of LIP and OP may be attributed to the single-center nature of our study (Shandong Province, China), which may be influenced by geographically specific immunophenotypes. Our study excluded patients with comorbid pulmonary hypertension (PH), which is often associated with severe fibrosis [40]. The relative proportion of inflammatory subtypes (LIP/OP) in our cohort may have been elevated after the exclusion of patients with PH.

The primary limitations of our study stem from the absence of several clinically relevant variables, including anti-U3RNP antibodies, anti-Th/To antibodies, RNA polymerase III antibodies, mRSS, telangiectasias, tendon friction rubs, and capillaroscopy. These variables are essential for a comprehensive assessment of SSc-ILD and were not included due to the constraints of our single-center retrospective data collection. While HRCT is a valuable tool for visualizing lung structure and identifying lesions, it has several limitations in the context of SSc-ILD. Specifically, HRCT may not be sufficiently sensitive to detect subtle changes in the lung interstitium that are critical for disease assessment. For instance, it may fail to capture early or mild interstitial changes, particularly in cases where the disease is in its nascent stages. Additionally, for patients requiring long-term follow-up, the cumulative radiation exposure from repeated HRCT examinations poses a potential risk to patient health. This concern is especially relevant given the need for frequent monitoring in progressive diseases like SSc-ILD. Future studies should integrate a broader range of clinical markers, including skin involvement and capillaroscopy, to enhance risk stratification and improve the comprehensiveness of disease assessment.

Conclusion

In conclusion, SSc-ILD is a clinically heterogeneous disease, beyond the known classic histologic classification (NSIP, UIP, OP, LIP et al.).Improved screening strategies can be used for early detection and intervention, better understanding of the mechanisms leading to SSc-ILD and identification of several future treatment options. With the paradigm shift in the diagnosis and treatment of SSc-ILD, we expect to improve methods to detect SSc phenotypes in order to better predict those patients who will develop ILD and to identify those who may respond to specific therapies.

Abbreviations

ILD	Interstitial lung disease
SSc	Systemic sclerosis
HRCT	High-resolution computed tomography
LIP	lymphocytic interstitial pneumonia
IIP	Idiopathic interstitial pneumonia
UIP	Usual interstitial pneumonia
NSIP	Nonspecific interstitial pneumonia
OP	Organized pneumonia
EULAR	European League Against Rheumatism
ACR	American College of Rheumatology
PFTs	Pulmonary function tests
PH	Pulmonary hypertension
GERD	Gastroesophageal Reflux Disease
RE	Reflux Esophagitis

SRC	Systemic sclerosis renal crisis
NLR	Neutrophil-to-leukocyte ratio
ELR	Eosinophil-to-leukocyte ratio
CAR	C-reactive protein–to-albumin ratio
RAR	Red blood cell distribution width–to-albumin ratio
PNI	Prognostic nutritional index
GGO	Ground glass opacity
IQR	Interquartile range
TCA	The two-step cluster analysis
RBCs	Red blood cells
KL-6	Lungen-6 antigen
MMF	Mycophenolate Mofetil
RA	Rheumatoid Arthritis
OSA	Obstructive sleep apnea
DM	Dermatomyositis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12890-025-03722-w.

Supplementary Material 1

Acknowledgements
We acknowledge the support of Shandong University and the Provincial Hospital of Shandong First Medical University.

Author contributions
Y.Z. and B.C. participated in the study design, literature review, and performed statistical analysis and presentation of the results and participated in the drafting and review of the manuscript. Y.Z., B. C., W.X., X.L., Z.J. and Z. Y. collected clinical data. Z.M. and Q. Y. participated in the study design, the drafting and review of the manuscript. All authors contributed to the article and approved the submitted version.

Funding
This work was supported by the National Natural Science Foundation of China (Youth fund project, Grant NO.82201994) and the Natural Science Foundation of Shandong Province (General Program, Grant NO. ZR2022MH016 and Youth fund project, Grant NO. ZR2021QH043) and the China Postdoctoral Science Foundation (grant 2023M742170).

Data availability
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Shandong Provincial Hospital, Shandong First Medical University (NSFC: NO.2022－413). All participants signed the informed consent form.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]The First Clinical Medical College, Shandong University of Traditional Chinese Medicine, Jinan, China
[2]Department of Rheumatology and Immunology, Cheeloo College of Medicine, Shandong Provincial Hospital, Shandong University, Jinan, Shandong, China
[3]Department of Rheumatology and Immunology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China
[4]Department of Radiology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

Received: 23 January 2025 / Accepted: 12 May 2025
Published online: 21 May 2025

References

1. Argula RG, Ward C, Feghali-Bostwick C. Therapeutic Challenges And Advances In The Management Of Systemic Sclerosis-Related Pulmonary Arterial Hypertension (SSc-PAH). Ther Clin Risk Manag. 2019;15:1427–1442. https://doi.org/10.2147/TCRM.S219024. PMID: 31853179; PMCID: PMC6916691.
2. Le Pavec J, Launay D, Mathai SC et al. Scleroderma lung disease. Clin Rev Allergy Immunol. 2011;40(2):104– 16. https://doi.org/10.1007/s12016-009-8194-2. PMID: 20063208.
3. 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(12):2069–73. https://doi.org/10.1136/annrheumdis-2014-205637. Epub 2014 Sep 30. PMID: 25269829.
4. Das A, Kumar A, Arrossi AV, Ghosh S, Highland KB. Scleroderma-related interstitial lung disease: principles of management. Expert Rev Respir Med. 2019;13(4):357–367. doi: 10.1080/17476348.2019.1575732. Epub 2019 Feb 11. PMID: 30686069.
5. Jee AS, Sheehy R, Hopkins P, Corte TJ, Grainge C, Troy LK, Symons K, Spencer LM, Reynolds PN, Chapman S, de Boer S, Reddy T, Holland AE, Chambers DC, Glaspole IN, Jo HE, Bleasel JF, Wrobel JP, Dowman L, Parker MJS, Wilsher ML, Goh NSL, Moodley Y, Keir GJ. Diagnosis and management of connective tissue disease-associated interstitial lung disease in Australia and new Zealand: A position statement from the thoracic society of Australia and new Zealand. Respirology. 2021;26(1):23–51. https://doi.org/10.1111/resp.13977. Epub 2020 Nov 24. PMID: 33233015; PMCID: PMC7894187.
6. Spagnolo P, Ryerson CJ, Putman R et al. Early diagnosis of fibrotic interstitial lung disease: challenges and opportunities. Lancet Respir Med. 2021;9(9):1065–1076. doi: 10.1016/S2213-2600(21)00017-5. Epub 2021 Jul 28. PMID: 34331867.
7. Chung JH, Montner SM, Adegunsoye A, Lee C, Oldham JM, Husain AN, MacMahon H, Noth I, Vij R, Strek ME. CT findings, Radiologic-Pathologic correlation, and imaging predictors of survival for patients with interstitial pneumonia with autoimmune features. AJR Am J Roentgenol. 2017;208(6):1229–36. Epub 2017 Mar 28. PMID: 28350485; PMCID: PMC6536259.
8. Valenzi E, Cody B, Lafyatis R. Usual interstitial pneumonia is the predominant histopathology in patients with systemic sclerosis receiving a lung transplant. Clin Exp Rheumatol. 2023;41(8):1670–8. https://doi.org/10.55563/clinexprheumatol/icr6hy. Epub 2023 Jun 29. PMID: 37382449; PMCID: PMC10528864.
9. Matson SM, Baqir M, Moua T, Marll M, Kent J, Iannazzo NS, Boente RD, Donatelli JM, Dai J, Diaz FJ, Demoruelle MK, Hamblin MB, Mathai SK, Ryu JH, Pope K, Walker CM, Lee JS. Treatment outcomes for rheumatoid Arthritis-Associated interstitial lung disease: A Real-World, multisite study of the impact of immunosuppression on pulmonary function trajectory. Chest. 2023;163(4):861–9. Epub 2022 Dec 5. PMID: 36470416; PMCID: PMC10107057.
10. Joven BE, Escribano P, Andreu JL, et al. 2013 ACR/EULAR systemic sclerosis classification criteria in patients with associated pulmonary arterial hypertension. Semin Arthritis Rheum. 2018;47(6):870–6. https://doi.org/10.1016/j.semarthrit.2017.10.006. Epub 2017 Oct 13. PMID: 29126717.
11. Sangani RA, Lui JK, Gillmeyer KR, Trojanowski MA, Bujor AM, LaValley MP, Klings ES. Clinical characteristics and outcomes in pulmonary manifestations of systemic sclerosis: contribution from pulmonary hypertension and interstitial lung disease severity. Pulm Circ. 2022;12(4):e12117. https://doi.org/10.1002/pul2.12117. PMID: 36238967; PMCID: PMC9535436.
12. Bruni C, Frech T, Manetti M. Vascular leaking, a pivotal and early pathogenetic event in systemic sclerosis: should the door be closed?? Front Immunol. 2018;9:2045. https://doi.org/10.3389/fimmu.2018.02045. PMID: 30245695; PMCID: PMC6137210.
13. Cole A, Ong VH, Denton CP. Renal disease and systemic sclerosis: an update on scleroderma renal crisis. Clin Rev Allergy Immunol. 2023;64(3):378–91. https://doi.org/10.1007/s12016-022-08945-x. Epub 2022 Jun 1. PMID: 35648373; PMCID: PMC10167155.
14. Khedoe P, Marges E, Hiemstra P, Ninaber M, Geelhoed M. Interstitial Lung Disease in Patients With Systemic Sclerosis: Toward Personalized-Medicine-Based Prediction and Drug Screening Models of Systemic Sclerosis-Related Interstitial Lung Disease (SSc-ILD). Front Immunol. 2020;11:1990. https://doi.org/10.3389/fimmu.2020.01990. PMID: 33013852; PMCID: PMC7500178.

15. Fu H, Zheng Z, Zhang Z, Yang Y, Cui J, Wang Z, Xue J, Chi S, Cao M, Chen J. Prediction of progressive pulmonary fibrosis in patients with anti-synthetase syndrome-associated interstitial lung disease. Clin Rheumatol. 2023;42(7):1917–1929. doi: 10.1007/s10067-023-06570-3. Epub 2023 Mar 17. Erratum in: Clin Rheumatol. 2023;42(7):1997–1998. https://doi.org/10.1007/s10067-023-06583-y. PMID: 36929316; PMCID: PMC10266998.
16. Kuwana M, Avouac J, Hoffmann-Vold AM, Smith V, Toenges G, Alves M, Distler O. Development of a multivariable prediction model for progression of systemic sclerosis-associated interstitial lung disease. RMD Open. 2024;10(3):e004240. https://doi.org/10.1136/rmdopen-2024-004240. PMID: 39242112; PMCID: PMC11381690.
17. Tibana RCC, Soares MR, Storrer KM, de Souza Portes Meirelles G, Hidemi Nishiyama K, Missrie I, Coletta ENAM, Ferreira RG, de Castro Pereira CA. Clinical diagnosis of patients subjected to surgical lung biopsy with a probable usual interstitial pneumonia pattern on high-resolution computed tomography. BMC Pulm Med. 2020;20(1):299. https://doi.org/10.1186/s12890-020-01339-9. PMID: 33198708; PMCID: PMC7670778.
18. Ramahi A, Lescoat A, Roofeh D, et al. Risk factors for lung function decline in systemic sclerosis-associated interstitial lung disease in a large single-centre cohort. Rheumatology (Oxford). 2023;62(7):2501–9. https://doi.org/10.1093/rheumatology/keac639. PMID: 36377780; PMCID: PMC10321078.
19. Lynch DA. Lung disease related to collagen vascular disease. J Thorac Imaging. 2009;24(4):299–309. https://doi.org/10.1097/RTI.0b013e3181c1acec. PMID: 19935226.
20. Belloli EA, Beckford R, Hadley R et al. Idiopathic non-specific interstitial pneumonia. Respirology. 2016;21(2):259– 68. doi: 10.1111/resp.12674. Epub 2015 Nov 13. PMID: 26564810.
21. Lee WJ, Kawahashi M, Hirahara H. Experimental analysis of Pendelluft flow generated by HFOV in a human airway model. Physiol Meas. 2006;27(8):661–74. https://doi.org/10.1088/0967-3334/27/8/001. Epub 2006 May 10. PMID: 16772665.
22. Glaspole I, Goh NS. Differentiating between IPF and NSIP. Chron Respir Dis. 2010;7(3):187– 95. https://doi.org/10.1177/1479972310376205. PMID: 20688895.
23. Wells AU, Cullinan P, Hansell DM et al. Fibrosing alveolitis associated with systemic sclerosis has a better prognosis than lone cryptogenic fibrosing alveolitis. Am J Respir Crit Care Med. 1994;149(6):1583-90. https://doi.org/10.1164/ajrccm.149.6.8004317. PMID: 8004317.
24. Sulku İ, Özer Gökaslan Ç. Role of computed tomography findings, complete blood count parameters and systemic inflammatory markers for predicting the severity in interstitial lung diseases. Sarcoidosis Vasc Diffuse Lung Dis. 2022;39(4):e2022042. https://doi.org/10.36141/svdld.v39i4.11616. PMID: 36533606; PMCID: PMC9798335.
25. Roman D, Iurciuc S, Caraba A. Pulmonary involvement in Sjögren's syndrome: correlations with biomarkers of activity and High-Resolution computer tomography findings. J Clin Med. 2024;13(4):1100. https://doi.org/10.3390/jcm13041100. PMID: 38398414; PMCID: PMC10889824.
26. Pugashetti JV, Lee JS. Overview of rheumatoid Arthritis-Associated interstitial lung disease and its treatment. Semin Respir Crit Care Med. 2024;45(3):329–41. https://doi.org/10.1055/s-0044-1782218. Epub 2024 Mar 14. PMID: 38484788.
27. Pihtili A, Bingol Z, Kiyan E, et al. Obstructive sleep apnea is common in patients with interstitial lung disease. Sleep Breath. 2013;17(4):1281–8. https://doi.org/10.1007/s11325-013-0834-3. Epub 2013 Apr 8. PMID: 23563999.
28. Tagaito Y, Isono S, Remmers JE, et al. Lung volume and collapsibility of the passive pharynx in patients with sleep-disordered breathing. J Appl Physiol (1985). 2007;103(4):1379–85. https://doi.org/10.1152/japplphysiol.00026.2007. Epub 2007 Jun 28. PMID: 17600160.
29. Zammit C, Liddicoat H, Moonsie I et al. Obesity and respiratory diseases. Int J Gen Med. 2010;3:335–43. doi: 10.2147/IJGM.S11926. PMID: 21116339; PMCID: PMC2990395.
30. Sarac S, Kavas M, Sahin M, et al. Relation of Warrick score and polysomnographic parameters in patients with interstitial lung disease. Med Sci Monit. 2019;25:2087–95. https://doi.org/10.12659/MSM.914905. PMID: 30894506; PMCID: PMC6439937.
31. Weng L, Chen Y, Liang T, et al. Biomarkers of interstitial lung disease associated with primary Sjögren's syndrome. Eur J Med Res. 2022;27(1):199. https://doi.org/10.1186/s40001-022-00828-3. PMID: 36217184; PMCID: PMC9549683.
32. Wang C, Yang H, Li T, et al. Analysis of characteristics related to interstitial lung disease or pulmonary hypertension in patients with dermatomyositis. Clin Respir J. 2023;17(12):1328–40. https://doi.org/10.1111/crj.13720. Epub 2023 Nov 20. PMID: 37985458; PMCID: PMC10730463.
33. Salonen J, Lampela H, Keskitalo E, et al. Bronchoalveolar lavage differential cell count on prognostic assessment of patients with stable or acute interstitial lung disease: A retrospective real-life study. Clin Immunol. 2020;220:108594. https://doi.org/10.1016/j.clim.2020.108594. Epub 2020 Sep 12. PMID: 32927080.
34. Cohen M, Giladi A, Gorki AD et al. Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. Cell. 2018;175(4):1031–1044.e18. https://doi.org/10.1016/j.cell.2018.09.009. Epub 2018 Oct 11. PMID: 30318149.
35. Elhai M, Sritharan N, Boubaya M, Balbir-Gurman A, Siegert E, Hachulla E, de Vries-Bouwstra J, Riemekasten G, Distler JHW, Rosato E, Del Galdo F, Mendoza FA, Furst DE, de la Puente C, Hoffmann-Vold AM, Gabrielli A, Distler O, Bloch-Queyrat C, Allanore Y. EUSTAR collaborators. Stratification in systemic sclerosis according to autoantibody status versus skin involvement: a study of the prospective EUSTAR cohort. Lancet Rheumatol. 2022;4(11):e785-e794. https://doi.org/10.1016/S2665-9913(22)00217-X. PMID: 38265945.
36. Santos CS, Morales CM, Castro CÁ, Álvarez ED. Clinical phenotype in scleroderma patients based on autoantibodies. Rheumatol Adv Pract. 2023;7(Suppl 1):i26–33. https://doi.org/10.1093/rap/rkad010. PMID: 36968636; PMCID: PMC10036993.
37. Bruni C, Tofani L, Fretheim H, Liem S, Velauthapillai A, Bjørkekjær HJ, et al. POS0388 DEVELOPING A SCREENING TOOL FOR THE DETECTION OF INTERSTITIAL LUNG DISEASE IN SYSTEMIC SCLEROSIS: THE ILD-RISC RISK SCORE. Annals of the Rheumatic Diseases. 2022;81(Suppl 1):449-450. doi: 10.1136/annrheumdis-2022-eular.2722. https://doi.org/10.1136/annrheumdis-2022-eular.2722
38. Hoffmann-Vold AM, Allanore Y, Alves M, Brunborg C, Airó P, Ananieva LP, Czirják L, Guiducci S, Hachulla E, Li M, Mihai C, Riemekasten G, Sfikakis PP, Kowal-Bielecka O, Riccardi A, Distler O. EUSTAR collaborators. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. Ann Rheum Dis. 2021;80(2):219–27. https://doi.org/10.1136/annrheumdis-2020-217455. Epub 2020 Sep 28. PMID: 32988845; PMCID: PMC7815627.
39. Manetti M. Correspondence on 'machine learning integration of scleroderma histology and gene expression identifies fibroblast polarisation as a hallmark of clinical severity and improvement'. Ann Rheum Dis. 2023;82(1):e21. https://doi.org/10.1136/annrheumdis-2020-219264. Epub 2020 Nov 6. PMID: 33158878.
40. Le Pavec J, Launay D, Mathai SC, Hassoun PM, Humbert M. Scleroderma lung disease. Clin Rev Allergy Immunol. 2011;40(2):104– 16. https://doi.org/10.1007/s12016-009-8194-2. PMID: 20063208.

Publisher's note