Roop *et al. Respiratory Research* (2026) 27:114
https://doi.org/10.1186/s12931-026-03536-2

Respiratory Research

RESEARCH Open Access

Artificial intelligence, machine learning-based automated fibrosis quantification in preclinical models of pulmonary fibrosis

Benjamin Roop[1†], Nery Matias Calmo[1,5†], Ashley Bass[1], Sarita Berigei[1], Rachel Knipe[1,3], Daniela Santos[1,3], Amalia DeCoursey[1], Gail Lee[4], Scott Turner[4], Markus Herrmann[2,3], Sreyankar Nandy[1,3†] and Lida Hariri[1,2,3,6*†]

Abstract

Background Precise, quantitative assessment of disease features on histological images from preclinical models is essential for therapeutic development in diseases such as pulmonary fibrosis. However, current histological fibrosis scoring methods, such as Ashcroft Scoring have several limitations, including high time and labor requirements by expert pathologist readers, subjective semi-quantitative assessment and interobserver variability. The goal of this study was to assess the feasibility of a supervised AI/ML-based framework for automated, rapid, objective quantification of fibrosis content and spatial distribution in histology slides obtained from a variety of preclinical fibrosis models.

Methods Lung histology slides stained with Masson's trichrome were obtained from 194 individual mice from two independent cohorts of preclinical mouse models of pulmonary fibrosis. A supervised AI/ML algorithm was trained, validated and independently tested to automatically detect, segment and quantify fibrosis compared against independent Ashcroft scoring by an expert pathologist reader. Spatial distribution of AI/ML-segmented fibrosis patterns were compared across histology images.

Results AI/ML-based fibrosis quantification demonstrated strong correlation with Ashcroft score, both in the validation cohort (Spearman $\rho = 0.85$, CI: 0.72–0.92), and in the independent, *de novo* test cohort (Spearman $\rho = 0.89$, CI: 0.84–0.93) with rapid assessment time (~ 1.5 times faster). Additionally, Ripley's K analysis revealed differences in spatial distribution of AI-segmented fibrosis patterns among samples with similar Ashcroft scores and overall fibrosis content.

Conclusions The AI/ML framework developed and independently validated in this study provides a robust, computationally-efficient method for precise, user-friendly, objective measurement of fibrosis content and spatial distribution, which would have major utility in preclinical therapeutic trials and investigations of disease pathogenesis.

†Benjamin Roop and Nery Matias Calmo contributed equally as first authors to this work. Sreyankar Nandy and Lida Hariri contributed equally as senior authors to this work.

*Correspondence:
Lida Hariri
lhariri@mgh.harvard.edu

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

Keywords Artificial intelligence, Machine learning, Ashcroft score, Bleomycin, Pulmonary fibrosis

Introduction

Effective therapeutic development in pulmonary fibrosis requires robust, preclinical animal studies before advancing to human clinical trials, both to assess therapeutic efficacy and investigate underlying disease pathogenesis [1, 2]. The bleomycin mouse model is a well-established model of pulmonary fibrosis used in pharmacologic preclinical studies [3]. Current methods for quantitative fibrosis assessment (i.e., hydroxyproline assay) require destructive tissue homogenization, prohibiting spatial assessment of fibrosis distribution/pattern [4, 5]. Ashcroft scoring, the current accepted method for visual histologic fibrosis assessment, provides only ordinal data, requires considerable effort and expertise from a pathologist reader, and is prone to interobserver variability [6–9]. Recent advances in digital pathology, including whole slide imaging, provide opportunities to investigate efficient, objective methods for automated quantification using machine learning (ML) and artificial intelligence (AI), which may overcome limitations of manual, semi-quantitative Ashcroft scoring [10, 11]. Here, we train, validate and independently test an AI/ML-based framework for automated detection, segmentation and quantification of fibrosis in digitized trichrome-stained histology slides obtained from a variety of preclinical fibrosis models and compare AI/ML-derived measurements against independent Ashcroft scoring.

Materials and methods

Preclinical models

Lung histology slides from 194 individual mice (156 bleomycin-treated and 38 healthy control sham) were stained with Masson's trichrome (one slide per mouse, 5 µm thickness) obtained from two independent research groups, referred as Cohort 1 (n = 60) [12, 13] and Cohort 2 (n = 134, from Pliant Therapeutics). Cohort 1 consisted of bleomycin-treated and untreated mice with transgenic endothelial-specific deletion of Rho kinase (ROCK1/2) or Sphingosine-1-phosphate receptor 1 (S1PR1) or global deletion of Sphingosine kinase 1 (SPHK1), therapeutic intervention and/or sham-saline treatment. Cohort 2 consisted of healthy control and bleomycin-treated mice with therapeutic intervention and sham-saline treatment. Histology slides from each cohort were digitized on separate slide scanners (*400x* magnification). Detailed information including the dosage, treatment administration route and duration from treatment administration to collecting sample for both cohorts are provided in Table 1.

AI/ML-based framework

Digitized slides from 7 mice in Cohort 1 were used to train the AI/ML algorithm. Fibrosis, normal alveoli, bronchi, bronchovascular support collagen, blood, and non-tissue areas were manually annotated on training slides using FIJI (https://imagej.net/software/fiji/). A supervised AI/ML algorithm, based on Trainable Weka Segmentation framework, was trained using manual annotations as ground truth [14]. The ability of the AI/ML algorithm to measure fibrosis was subsequently validated on the remaining 53 mice from Cohort 1 and independently tested *de novo* on 134 mice from Cohort 2. Similar to Ashcroft scoring, histology artifacts were visually identified and excluded from analysis. Fibrosis percentage was automatically calculated by dividing the number of pixels labeled as 'fibrosis' by the AI/ML algorithm by the total number of pixels in the tissue for each slide/mouse.

Ashcroft scoring

A board-certified pathologist with expertise in Ashcroft scoring applied the modified Ashcroft scale to each histology slide as previously described, blinded to AI/ML algorithm results [6, 7]. Each slide was divided into 20x fields of view, which were each assigned an individual Ashcroft score and subsequently averaged to obtain a total Ashcroft score for each mouse.

Statistical analysis

Percent fibrosis quantified by the AI/ML algorithm was compared against averaged Ashcroft score for each mouse using Spearman's rank correlation for the validation dataset from Cohort 1 and separately for the *de novo* test dataset from Cohort 2. P-value of < 0.05 was considered statistically significant (R, version 4.0.2). Multi-distance Ripley's K analysis (Python, version 3.9) was performed to assess the spatial distribution of AI/ML-segmented fibrosis patterns from each histology image in both cohorts [15].

Results

Fibrosis percentage measured by the AI/ML framework for the validation dataset in Cohort 1 (53 mice) ranged 0.51–37.76% (mean: 8.62%, SD: 8.27%), while the averaged Ashcroft scores ranged 0.10–4.73 (mean: 2.35, SD: 1.44). AI/ML-based percent fibrosis measurement demonstrated strong correlation with Ashcroft score in Cohort 1 (Spearman $\rho = 0.85$, CI: 0.72–0.92). The AI/ML-based fibrosis measurement in Cohort 2 (134 mice) ranged 0.14–77.81% (mean: 18.12%, SD: 1.15%) and averaged Ashcroft scores ranged 0.27–5.54 (mean: 3.46, SD:

Table 1 Preclinical mouse models included in study cohorts 1 and 2. Cohort 1 ($n=60$ mice/slides) contains lung histology from experiments performed in the same research laboratory that were used for both training ($n=7$) and testing of the AI/ML model ($n=53$). Cohort 2 ($n=134$ mice/slides) contains lung histology from experiments performed in a separate research laboratory, prepared in a different histology facility, and scanned with a different digitized scanner manufacturer than cohort 1. Cohort 2 was used solely for *de novo* independent testing of the AI/ML model. * Indicates the number of slides from a group in cohort 1 used for model training. Parenthetical number indicates publication reference for relevant study

Treatment and/or Modification	Total Number	Treated with Bleomycin	Not Treated with Bleomycin	Bleomycin Dosage	Treatment Route	Harvest Day after Bleomycin	Digital Scanner
Cohort 1							
Endothelial S1PR1 deletion [12]	19	11*	8	1.2 or 0.5U/kg	Oral	0, 7, or 14 days	Hamamatsu Nanozoomer
Endothelial S1PR1 floxed control [12]	14	8	6				
Sphingosine-1 Kinase deletion	6	6***	0				
Endothelial ROCK1 deletion	3	3	0				
Endothelial ROCK1 floxed control	2	2	0				
Endothelial ROCK2 deletion	2	2	0				
Endothelial ROCK2 floxed control	1	1	0				
Simvastatin [13]	4	3**	1				
Saline [12, 13]	9	8*	1				
Cohort 2							
3G9	18	18	0	3 then 2 then 1.5U/kg	Oral	28 days	Leica Aperio
IgG	18	18	0				
Integrin Inhibitor 1	40	40	0				
Integrin Inhibitor 2	18	18	0				
Saline	18	18	0				
Water	22	0	22	N/A			

1.36). Spearman rank correlation between the AI/ML-based fibrosis measurement in Cohort 2 showed strong correlation with Ashcroft score (Spearman $\rho=0.89$, CI: 0.84–0.93), demonstrating accuracy of AI/ML-based fibrosis quantification in an independent, *de novo* test dataset. The AI/ML framework demonstrated high accuracy for histology images with patchy, minimal fibrosis content (Fig. 1a), high fibrosis content (Fig. 1b) and no fibrosis in control mice (Fig. 1c). Mean time required for assessment of one slide/mouse was 6.8 min for the AI/ML algorithm versus >10 min for Ashcroft scoring. Ripley's K analysis evaluating the spatial distribution of AI-segmented fibrosis patterns show clear differences amongst samples with varying fibrotic distributions, including between histology images with similar Ashcroft scores. Histology slides from mouse lungs with sparse, evenly distributed fibrosis shows an observed Ripley's K plot with high similarity to the predicted Ripley's K plot for completely random spatial distribution (Fig. 2a). Histology slides from mouse lungs with more dense, clustered fibrosis regions interspaced with unaffected lung parenchyma (Figs. 2b and c) show deviation from the predicted Ripley's K plot for completely random spatial distribution.

Discussion

Use of preclinical models of lung fibrosis in pathophysiologic and pharmacologic investigational studies require rapid, precise methods to objectively identify and quantify fibrosis content. Ashcroft scoring is time- and labor-intensive, semi-quantitative, requires expertise, and has known interobserver variability. Here, we developed, trained, and independently validated an AI/ML framework to automatically detect, segment and quantify fibrosis on histology, including in bleomycin-treated and untreated mice with transgenic modifications, therapeutic interventions, sham-saline treatment, and healthy control mice. The two cohorts consisted of different mouse populations, with transgenic mice prevalent in Cohort 1 and therapeutically-treated mice prevalent in the independent *de novo* test Cohort 2, with staining performed in different labs and slides scanned on different scanner systems/manufacturers. We demonstrate high accuracy of the AI/ML framework for fibrosis quantification in both Cohorts, highlighting the robustness and general applicability of the algorithm for use in preclinical studies of fibrosis.

The current AI/ML framework allowed for ~ 1.5 times faster assessment of fibrosis per mouse compared to Ashcroft scoring by an expert reader (6.8 versus > 10 min), demonstrating comparable accuracy while reducing time

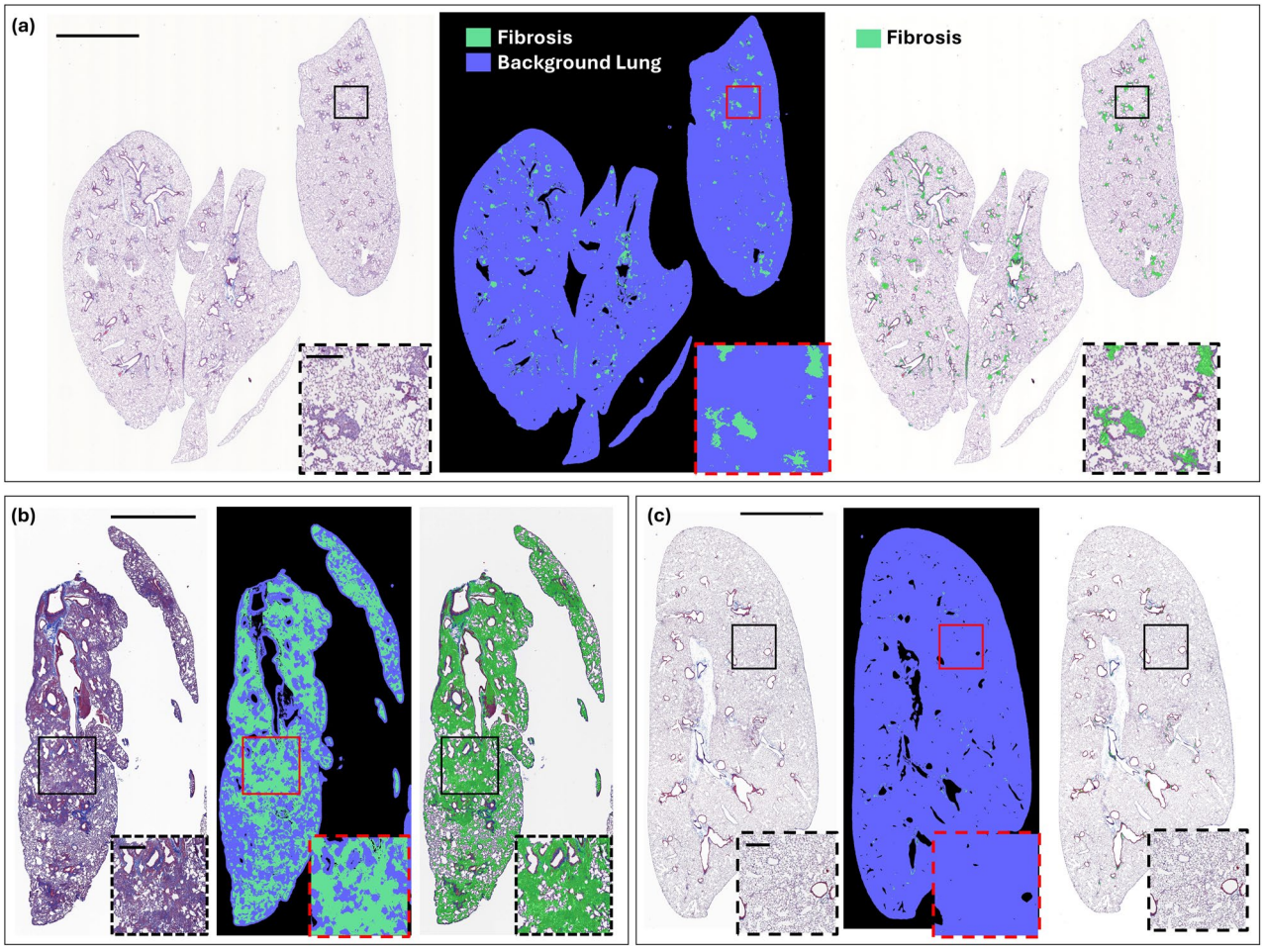

Fig. 1 AI/ML-based algorithm for fibrosis segmentation and quantification in preclinical mouse models of fibrosis. (**a**) Patchy, low fibrosis content and (**b**) high fibrosis content in preclinical mouse model of fibrosis and (**c**) no fibrosis content in healthy control mouse is accurately segmented by the AI/ML-based algorithm. Original histology scan (left), AI/ML-segmented fibrosis mask (green) overlaid on background lung (blue) (middle), and overlaid AI/ML-segmentation of fibrosis (green) on histology (right), (20x), with boxed regions magnified (bottom right inset, 100x). Scalebar on the main histology figures: 2.5 mm, Scalebar on the magnified inset figures: 250 µm

and labor required from an experienced pathologist. The discrete, subjective nature of Ashcroft scoring is subject to known interobserver variability. Automated, AI/ML-based quantification provides more precise, objective, continuous measurement of fibrosis. Previous studies using AI/ML to perform automatic Ashcroft scoring and fibrosis quantification showed similar results but had several limitations, including output of Ashcroft class probabilities rather than continuous measurements, use of specialized software not readily available, and lack of generalizability and robustness of their AI/ML algorithms across various preclinical models of fibrosis, scanner models, and histopathology staining laboratories [11, 16].

The objective fibrosis segmentation from the AI/ML framework provides unique opportunities to quantitatively assess fibrosis spatial distribution, in addition to measuring bulk fibrosis content. Representative examples of Ripley's K plots in Fig. 2 show the utility of the spatial analysis as it pertains to our AI/ML fibrosis quantification framework. In examples of lung histology slides with similar Ashcroft scores, the Ripley K plots showed differences between histology images with sparse, evenly distributed fibrosis and images with dense, clustered fibrosis regions interspaced with unaffected lung parenchyma. This highlights an added advantage of the AI/ML framework over classical Ashcroft scoring. Future studies may evaluate the role of spatial distribution analysis in the context of different mouse models and assessment of disease pathogenesis and treatment response.

There were some limitations in this study. The current framework used manual supervision to identify coverslip air bubbles, cutting, and/or staining artifacts and exclude them from analysis, similar to a pathologist reader. Future iterations may integrate automated analysis tools to detect and exclude artifacts [17]. Although

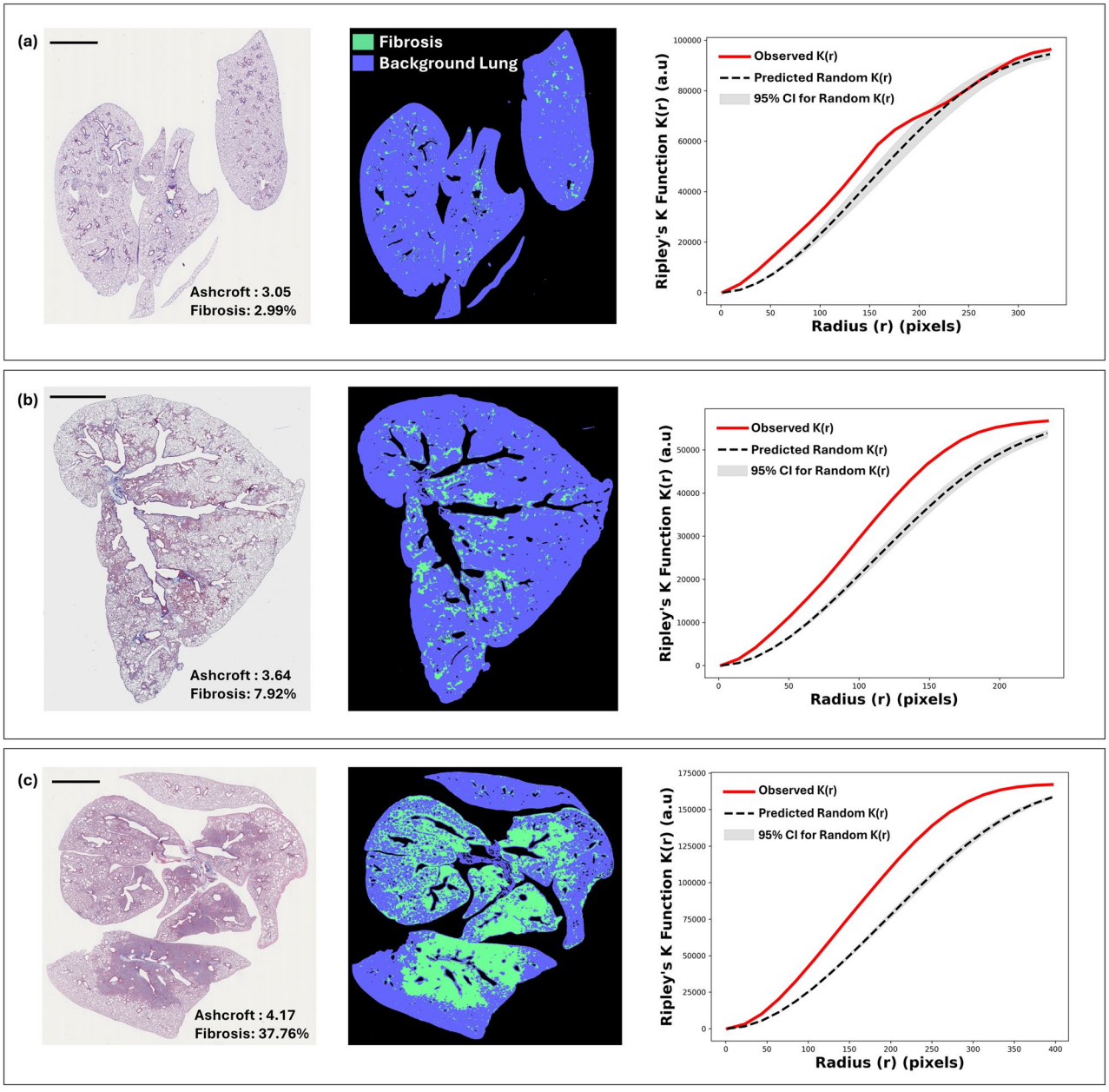

Fig. 2 Spatial analysis of fibrosis patterns segmented by the AI/ML-based algorithm. Ripley's K analysis evaluating the spatial distribution of AI/ML-segmented fibrosis patterns is plotted as a function of different search radius in pixels. The observed K-distribution (red line) is compared against a predicted K-distribution with completely random spatial distribution (black dashed line) to determine if the distribution is dispersed or clustered. Each panel shows the original histology image (left), the AI/ML-segmented fibrosis (green) overlaid on background lung (blue) (middle), and corresponding Ripley's K-distribution plots (right) showing the observed Ripley's K- distribution (red line), completely spatially random distribution (black dashed line) and 95% confidence intervals (gray). (**a**) In a sample with AI/ML-segmented sparse, evenly dispersed fibrosis pattern, the observed Ripley's K-distribution (red line) shows close proximity to the predicted spatially random distribution (black dashed line). (**b**) In a sample with similar Ashcroft score as (**a**) but with slightly higher AI/ML-segmented fibrosis content, the spatial distribution shows more densely aggregated fibrosis clusters with intervening regions of unaffected lung parenchyma. The corresponding observed Ripley's K-distribution (red line) shows deviation from the predicted random distribution (black dashed line). (**c**) Sample with high AI/ML-predicted fibrosis content with dense fibrosis clusters alternating with regions of unaffected lung parenchyma. The corresponding observed Ripley's K-distribution (red line) shows consistent deviation from the predicted spatially random distribution (black dashed line). Scalebar on the main histology figures: 2.5 mm

hydroxyproline assay is an accepted method of fibrosis quantification, it is destructive of the lung tissue by nature, and therefore, direct comparison cannot be made with the AI/ML-based segmentation model that requires histology images. A relatively small fraction of cases were manually annotated for model training, as compared to conventional larger AI/ML training datasets. Despite this, the AI/ML framework generalized well and demonstrated high accuracy on histology images from a wide variety of preclinical fibrosis models.

Conclusion

In conclusion, we successfully trained, validated, and independently tested an AI/ML-based framework for precise quantification of fibrosis on histology from preclinical mouse models. This provides a robust, computationally-efficient method for objective, reproducible measurement of fibrosis content and spatial distribution, which has utility for both therapeutic trials and investigations of disease pathogenesis using preclinical models.

Abbreviations

AI	Artificial intelligence
ML	Machine learning
ROCK	Rho-associated coiled-coil-forming protein kinase
SPHK	Sphingosine Kinase 1
SD	Standard deviation

Acknowledgements

The authors gratefully acknowledge Jenny Zhou, MD, PhD, and the Massachusetts General Hospital Wellman Center for Photomedicine Photopathology Core, for their assistance with digital slide scanning. We also acknowledge Susan Sheng, PhD, from the Division of Pulmonary and Critical Care Medicine at Massachusetts General Hospital for helping with review and revision of the manuscript.

Authors' contributions

BR, SN, NMC, MH, LH participated in the study design. RK, DS, GL, ST prepared and provided mice and histology slides. BR, NMC, AB, SB, AD, MH, SN, LH participated in algorithm development and validation. LH performed Ashcroft scoring. BR, SN, LH wrote the manuscript with input from all authors. All authors gave final approval of the manuscript.

Funding

This study was funded in part by the National Institutes of Health (Grant numbers.
K23HL132120, K25HL169846, K08HL140175, R01HL152075, R01HL169225).

Data availability

Data can be made available from the corresponding author on reasonable request. The codes to generate AI/ML-segmentation and fibrosis quantification are available in the GitHub repository (https://github.com/Nery-MC/Preclinical-Pulmonary-Fibrosis-WekaSegmentation).

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

LH reports grants from Boehringer Ingelheim and has received personal consulting fees from Boehringer Ingelheim, Pliant Therapeutics, Clario, Abbvie, Bristol Myers Squibb, and Biogen Idec. SN reports grant from Boehringer Ingelheim. Although Pliant Therapeutics provided mouse histology data, Pliant had no role in the design, analysis, or interpretation of the results of this study. MH is an employee and shareholder of and has intellectual property interests with F. Hoffmann-La Roche AG.

Author details

[1]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, MABoston, USA
[2]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
[3]Harvard Medical School, Boston, MA, USA
[4]PLIANT Therapeutics, South San Francisco, CA, USA
[5]Department of Biomedical Engineering, Tufts University, Medford, MA 02155, USA
[6]Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114, USA

Received: 26 October 2025 / Accepted: 21 January 2026
Published online: 04 February 2026

References

1. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of idiopathic pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med. 2018;198(5):e44–68.
2. Richeldi L, Varone F, Bergna M, de Andrade J, Falk J, Hallowell R, et al. Pharmacological management of progressive-fibrosing interstitial lung diseases: a review of the current evidence. Eur Respir Rev. 2018;27(150):180074.
3. Adamson IY, Bowden DH. The pathogenesis of bleomycin-induced pulmonary fibrosis in mice. Am J Pathol. 1974;77(2):185–97.
4. Reddy GK, Enwemeka CS. A simplified method for the analysis of hydroxyproline in biological tissues. Clin Biochem. 1996;29(3):225–9.
5. Sikic BI, Young DM, Mimnaugh EG, Gram TE. Quantification of bleomycin pulmonary toxicity in mice by changes in lung hydroxyproline content and morphometric histopathology. Cancer Res. 1978;38(3):787–92.
6. Ashcroft T, Simpson JM, Timbrell V. Simple method of estimating severity of pulmonary fibrosis on a numerical scale. J Clin Pathol. 1988;41(4):467–70.
7. Hubner RH, Gitter W, El Mokhtari NE, Mathiak M, Both M, Bolte H, et al. Standardized quantification of pulmonary fibrosis in histological samples. Biotechniques. 2008;44(4):507–11. 14 – 7.
8. Dettori JR, Norvell DC. The anatomy of data. Global Spine J. 2018;8(3):311–3.
9. Simler NR, Howell DC, Marshall RP, Goldsack NR, Hasleton PS, Laurent GJ, et al. The Rapamycin analogue SDZ RAD attenuates bleomycin-induced pulmonary fibrosis in rats. Eur Respir J. 2002;19(6):1124–7.
10. Bordag N, Biasin V, Schnoegl D, Valzano F, Jandl K, Nagy BM, et al. Machine learning analysis of the bleomycin mouse model reveals the compartmental and Temporal inflammatory pulmonary fingerprint. iScience. 2020;23(12):101819.
11. Seger S, Stritt M, Vezzali E, Nayler O, Hess P, Groenen PMA, et al. A fully automated image analysis method to quantify lung fibrosis in the bleomycin-induced rat model. PLoS ONE. 2018;13(3):e0193057.
12. Knipe RS, Spinney JJ, Abe EA, Probst CK, Franklin A, Logue A, et al. Endothelial-Specific loss of Sphingosine-1-Phosphate receptor 1 increases vascular permeability and exacerbates Bleomycin-induced pulmonary fibrosis. Am J Respir Cell Mol Biol. 2022;66(1):38–52.
13. Santos DM, Pantano L, Pronzati G, Grasberger P, Probst CK, Black KE, et al. Screening for YAP inhibitors identifies Statins as modulators of fibrosis. Am J Respir Cell Mol Biol. 2020;62(4):479–92.
14. Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW, Schindelin J, Cardona A, et al. Trainable Weka segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics. 2017;33(15):2424–6.
15. Behanova A, Klemm A, Wahlby C. Spatial statistics for Understanding tissue organization. Front Physiol. 2022;13:832417.
16. Heinemann F, Birk G, Schoenberger T, Stierstorfer B. Deep neural network based histological scoring of lung fibrosis and inflammation in the mouse model system. PLoS ONE. 2018;13(8):e0202708.

17. Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: an Open-Source quality control tool for digital pathology slides. JCO Clin Cancer Inf. 2019;3:1–7.

Publisher's Note