La radiologia medica (2026) 131:634–647
https://doi.org/10.1007/s11547-025-02166-w

CHEST RADIOLOGY

Advanced lung segmentation on chest HRCT: comprehensive pipeline for quantification of airways, vessels, and injury patterns

Alberto Arrigoni[1,2] · Francesca Pennati[2] · Pietro Andrea Bonaffini[3] · Alberto Senatieri[3] · Gregorio Chierchia[3] · Chiara Allegri[4] · Caterina Conti[4] · Fabiano Di Marco[4] · Anna Caroli[1] · Andrea Aliverti[2]

Received: 9 June 2025 / Accepted: 2 December 2025 / Published online: 24 December 2025

Abstract

Purpose Chest high-resolution computed tomography (HRCT) is crucial for diagnosing and monitoring pulmonary diseases involving parenchymal, vascular, and airway alterations. However, segmentation faces challenges in distinguishing pulmonary structures due to heterogeneity in image acquisition and pathological manifestations. Unlike existing tools, which usually target a single anatomical structure and rely predominantly on either deep learning or rule-based approaches, our hybrid pipeline pairs U-Net-based AI segmentation with tailored image processing refinements to produce a reliable and simultaneous segmentation of lungs, airways, pulmonary vessels, and parenchymal injury patterns, while enabling quantitative characterization across a spectrum of disease severities and types (inflammatory and infectious).

Methods This retrospective observational study employed 19 chest CT scans from COVID-19 public datasets for deep learning, 8 annotated scans from the EXACT'09 challenge to validate airway segmentation, and 20 retrospective HRCT scans from COVID-19 and idiopathic pulmonary fibrosis patients for pipeline validation. The pipeline performs preliminary segmentation of lungs, airways, and pathological regions using U-Nets, followed by image processing to refine results, include vasculature, and classify injury patterns in ground-glass opacities, reticulations/consolidations, and air-filled pathological spaces. Three radiologists validated segmentations on a 1–5 scale, and the Kruskal–Wallis test was conducted to assess differences across raters, pathologies, and severities.

Results The proposed pipeline visually outperformed established tools (LungCTAnalyzer, PTK, TotalSegmentator). Airway's segmentation achieved a Dice coefficient of 0.91 [0.89–0.92] on the EXACT'09 dataset. Radiologists assigned scores of 4 and 5 to segmentation completeness and accuracy, respectively, for both airways and vessels. Parenchymal injury patterns

Anna Caroli and Andrea Aliverti have equally contributed for this work.

✉ Anna Caroli
anna.caroli@marionegri.it

Alberto Arrigoni
alberto.arrigoni@marionegri.it

Francesca Pennati
francesca.pennati@polimi.it

Pietro Andrea Bonaffini
pbonaffini@asst-pg23.it

Alberto Senatieri
a.senatieri@campus.unimib.it

Gregorio Chierchia
g.chierchia@campus.unimib.it

Chiara Allegri
callegri@asst-pg23.it

Caterina Conti
caterina.conti1@gmail.com

Fabiano Di Marco
fdimarco@asst-pg23.it

Andrea Aliverti
andrea.aliverti@polimi.it

[1] Department of Biomedical Engineering, Istituto Di Ricerche Farmacologiche Mario Negri IRCCS, Villa Camozzi Via G.B. Camozzi 3, 24020 Ranica, BG, Italy

[2] Dipartimento Di Elettronica, Informazione E Bioingegneria, Politecnico Di Milano, Via Ponzio 34/5, 20133 Milan, MI, Italy

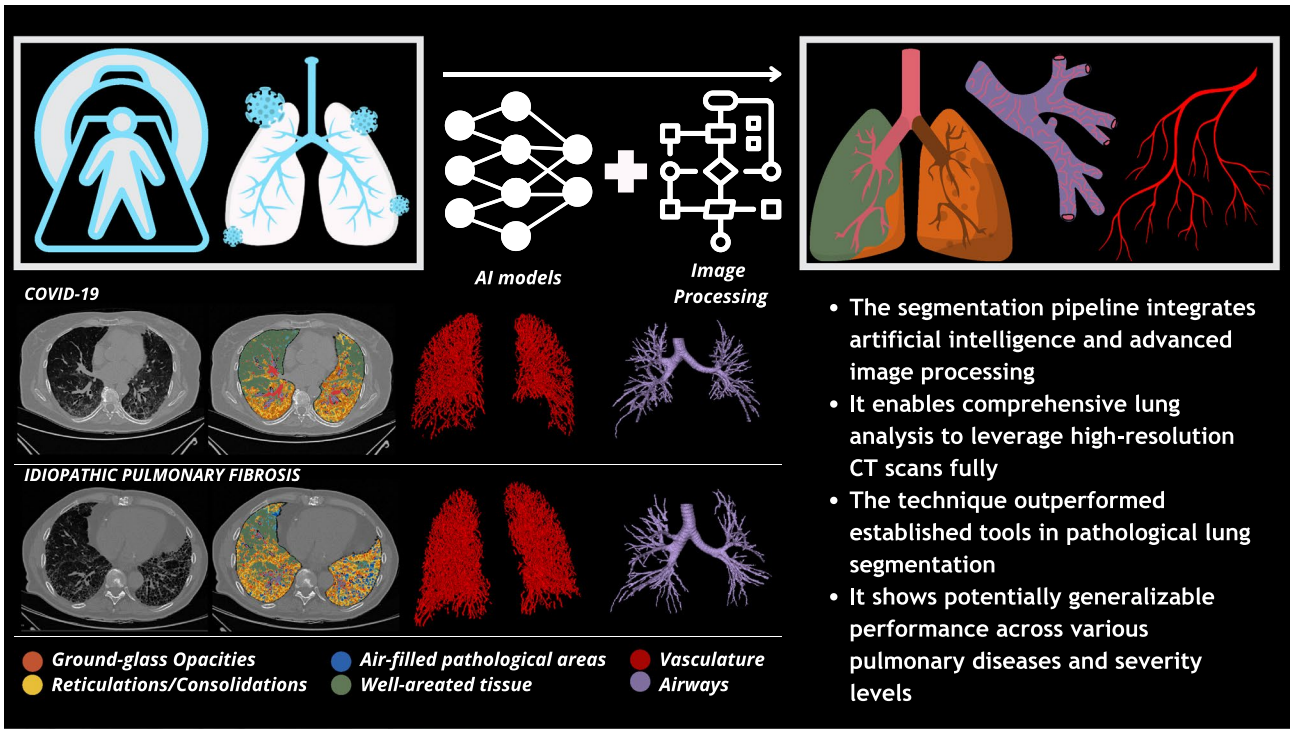
[3] Department of Radiology, ASST Papa Giovanni XXIII, Piazza OMS 1, 24127 Bergamo, BG, Italy

[4] Respiratory Unit, ASST Papa Giovanni XXIII, Piazza OMS 1, 24127 Bergamo, BG, Italy

scored 4 for completeness, accuracy, and classification. Ratings were consistently high with no significant differences among raters, diseases, and severity levels.

Conclusion The proposed pipeline introduces a novel, comprehensive, and hybrid approach for simultaneous, multi-structure lung segmentation, demonstrating reliable and potentially generalizable performance across inflammatory and infectious pulmonary diseases.

Graphical Abstract



Keywords Lung · Chest Radiology · Computed Tomography · Segmentation · Image Processing · Artificial Intelligence

Abbreviations

CIP	Chest imaging platform
CNNs	Convolutional neural networks
FA	Fractional anisotropy
GGO	Ground-glass opacities
HRCT	High-resolution CT
ILD	Interstitial lung disease
IoU	Intersection over union
IPF	Idiopathic pulmonary fibrosis
PTK	Pulmonary toolkit

Introduction

High-resolution CT (HRCT) is essential for diagnosing and monitoring pulmonary diseases, revealing subtle lung parenchymal changes. It detects increased attenuation areas, such as ground-glass opacities (GGO), consolidations, reticulations, and honeycombing walls [1, 2], as well as decreased attenuation areas, including emphysema, bullae, and cysts—indicative of abnormal air-filled spaces lacking alveolar architecture [3–5]. Additionally, HRCT provides insights into the airways and pulmonary vasculature, highlighting bronchial dilation (bronchiectasis) or stenosis [3–6] and modifications in the size, shape, and distribution of pulmonary vessels [7]. Recent studies have underscored the key role of HRCT in assessing COVID-19 severity and post-acute sequelae [3, 4, 8], as well as diagnosing diseases such as interstitial lung disease (ILD), including idiopathic pulmonary fibrosis (IPF) [9–11].

Despite its potential, HRCT evaluation in clinical practice mostly relies on visual assessment without any quantitative characterization due to the lack of accurate, robust, and generalizable segmentation techniques.

Manual methods are labor-intensive and prone to inter-observer variability. Automatic methods, traditionally reliant on intensity thresholds, morphological operations, and filtering, struggle with intensity similarities between normal

pulmonary structures and abnormalities (e.g., vessels vs. dense injury patterns, airways vs. cysts) and variability in image acquisition parameters and pathological manifestations. Deep learning, with convolutional neural networks (CNNs), offers promising solutions [12–15] but also faces difficulties with diverse pathological features, particularly if not trained on datasets sufficiently representative of the wide variability of possible disease facets and severity stages. Therefore, available techniques perform naïve segmentations and focus on specific lung structures. The Pulmonary Toolkit (PTK) and the Chest Imaging Platform (CIP) are widely used for lung segmentation from CT scans. PTK (https://github.com/tomdoel/pulmonarytoolkit) employs region growing, morphological operations, and vessel enhancement but performs poorly in complex cases. Similarly, CIP (https://chestimagingplatform.org/) and its LungCTAnalyzer extension use intensity-based segmentation and region growing, often leading to inaccuracies. In the realm of AI-based tools, the TotalSegmentator model (https://github.com/wasserth/TotalSegmentator), based on the U-Net [16] and nnU-Net [17] (https://github.com/MIC-DKFZ/nnUNet), shows promise but struggles with severely altered images.

Recent progress in pulmonary CT segmentation has leveraged 3D Transformer-based architectures.

[15, 18, 19] and semi-supervised learning [20–22] to achieve high accuracy for specific structures such as nodules, vessels, and tumors. Several approaches have also addressed parenchymal injury segmentation [23–25]. However, despite these advances, current methods are limited: they only segment single or paired anatomical structures (airways and/or vessels) [26, 27] or address pathological changes separately, and do not offer integrated multi-structure segmentation.

The comprehensive, simultaneous segmentation of both anatomical structures and diverse parenchymal injury patterns across different pulmonary diseases remains an unmet need in the field, where a major challenge is the sizeable anatomical variability and lung disease heterogeneity.

To the best of our knowledge, no available tools provide a precise and comprehensive assessment of lung architecture with the capability to accurately differentiate between different structures in the case of pathological pictures.

To answer this unmet need, we introduce, test, and validate a novel segmentation pipeline that combines traditional image processing and deep learning techniques to ensure a quantitative assessment of lung architecture—encompassing vasculature, airways, and injury patterns—on chest HRCT scans.

Methods

Datasets

For the development and evaluation of this study, we considered retrospective CT data from different sources to ensure an adequate representation of various scanning protocols, scanner types, image resolutions, and heterogeneity of pathological abnormalities. Precisely, we used 19 non-contrast enhanced chest CT scans from patients with COVID-19, sourced from the Radiopaedia (https://www.kaggle.com/competitions/covid-segmentation/data) and Coronacases dataset (https://coronacases.org/), to train a U-Net segmenting the pathological tissue with increased attenuation; 8 CT scans from the EXACT'09 dataset (http://image.diku.dk/exact/) to quantitatively validate the airway segmentation; and 20 retrospective HRCT scans acquired at the ASST Papa Giovanni XXIII Hospital (Bergamo, Italy) from patients with COVID-19 and IPF to validate the proposed pipeline. These latter scans were obtained as part of two observational studies approved by the local Research Ethics Committee (Registration No. 2023/0233 and Registration No. 2023/0047).

For the IPF cohort, we included all patients with technically valid HRCT studies (i.e., no severe motion artifact, complete lung coverage, and adequate inspiratory level). A radiologist with more than 10 years of experience grouped images into "mild" and "severe" examples according to the extent of the radiological impairment. For the COVID-19 cohort, scans were first grouped into the two severity classes, and a stratified random sampling was performed to ensure the selection of 10 mild and 10 severe cases overall. No healthy control group was included, as the primary objective of the technical validation was to evaluate the robustness of the pipeline for segmentation in heterogeneous, clinically relevant disease patterns. However, very mild COVID-19 cases with only localized and minor abnormalities were included to sample the lower end of pathological burden and approximate healthy situations.

Datasets details are provided in the Supplementary material (Supplementary Table 1 and Supplementary Fig. 1).

All scans were obtained in the supine position at full inspiration. Scans were converted from DICOM to Nifti format and resampled on the axial plane to 512×512 pixels using a linear interpolator.

HRCT processing pipeline

The chest HRCT processing pipeline proposed in this study was developed using Bash (v5.2.15, https://www.

gnu.org/software/bash/), MATLAB (vR2021a, https://www.mathworks.com/), and Python (v3.7.10; https://www.python.org/), exploiting pertinent libraries, including TensorFlow (https://www.tensorflow.org/) for AI developments and the MATLAB Image Processing Toolbox for image processing and analysis.

The pipeline takes chest HRCT scans as input and first performs lung segmentation and parcellation into anatomical lobes, plane-oriented segments, and depth-based layers. Then, it segments the pulmonary airways and vessels, categorizing the latter by dimension. For pathological conditions altering the parenchyma structure, the algorithm differentiates between well-aerated and abnormal tissue, classifying the latter into three injury patterns based on the CT attenuation of the alteration.

AI techniques—particularly U-Net models, are used for preliminary segmentations, followed by extensive post-processing or complementary segmentation.

Representative examples of lung segmentation and parcellation are shown in Supplementary Fig. 2. The preliminary lung and lobe segmentation is obtained using the publicly available U-Net trained models, R231 and LTRCLobes [14]. Then, each lung is divided into two sections along the transversal and sagittal axes and three along the longitudinal axis. Last, to obtain a distance-based parcellation, the convex hull merging the two lungs by the mediastinum is performed, and the Euclidean distance transform is applied to obtain a distance map, enabling the labeling of distal, middle, and inner lung regions.

The airway segmentation pipeline is outlined in Fig. 1 and Supplementary Fig. 3 and includes a first naïve segmentation with a U-Net model—BronchiNet (https://github.com/antonioguj/bronchinet) [13]—and post-processing aimed at removing false positives and integrating missing branches.

In particular, false positives are detected in thick structures in the distal layer, where airways normally narrow with depth, and in smaller objects that do not exhibit an elongated, tubular morphology.

The trachea and main bronchi are precisely included in the segmentation using a morphological reconstruction technique [28]. In this process, seeds corresponding to clearly identified tracheal sections are grown within the mask of the low-attenuation regions confined to the mediastinum. In addition, missing branches deep in the lungs are incorporated by applying the Jerman 3D vessel enhancement [29] filter to the low-attenuation mask inside the lungs and selecting tubular structures originating at the extremes of current airway segmentation.

Following this refinement, the airway structure is divided into two anatomical components: the trachea and the branches. The division is determined at the carina, where the centerline bifurcates into the primary bronchi, ensuring a standardized anatomical separation.

The segmentation and classification of injury patterns is outlined in Fig. 2 and Supplementary Fig. 4. Specifically, the algorithm detects patterns of increased lung attenuation, differentiating between GGO and reticulations/consolidations, and identifies patterns of decreased lung attenuation, which accounts for conditions such as emphysema, bullae, and air cysts.

High attenuation regions in the lung parenchyma are initially segmented using an in-house model trained on the Radiopaedia and Coronacases datasets.

Specifically, we employed a U-Net architecture with an EfficientNetB0 backbone initialized with pre-trained weights from ImageNet (https://www.image-net.org). To accommodate the single-channel CT images, an additional Conv2D layer with a 1×1 kernel was introduced to map them into the 3-channel format. The model produced a

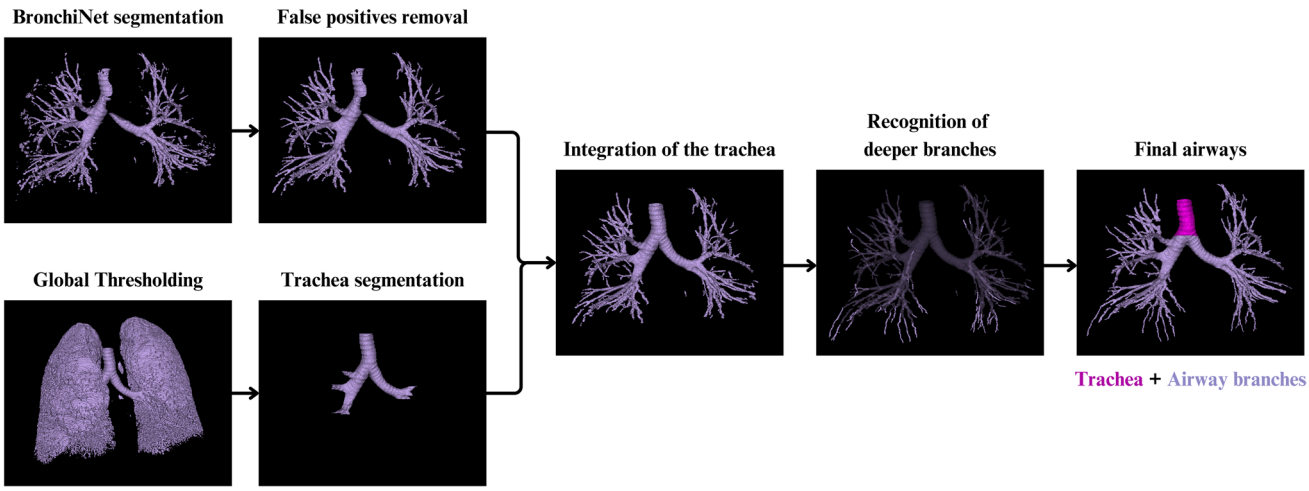


Fig. 1 Schematic representation of the airway segmentation algorithm

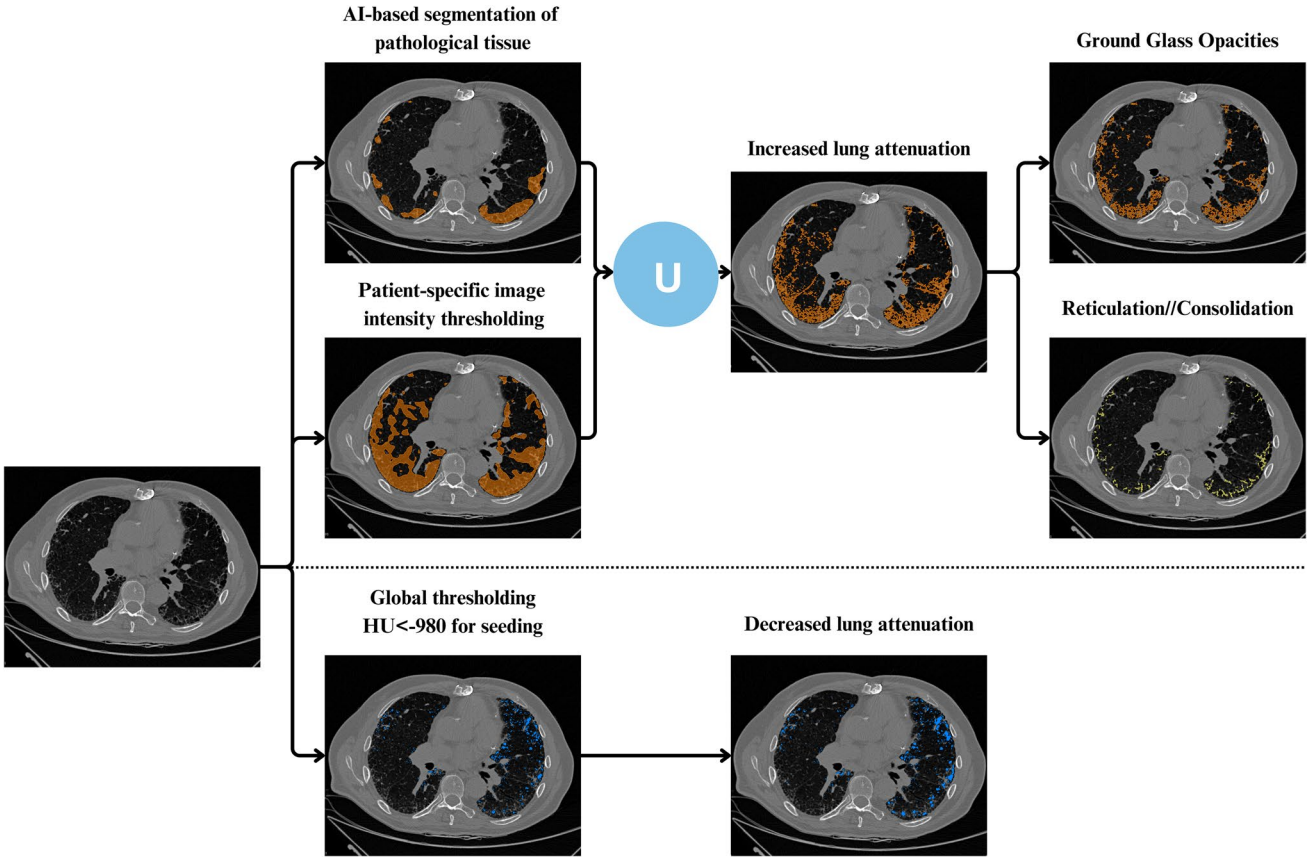


Fig. 2 Schematic representation of the algorithm for segmentation and classification of lung disease patterns

single-class segmentation output with a sigmoid activation function. Training was configured with the Adam optimizer (learning rate = 0.0001), a batch size of 9, and a total of 32 initial epochs followed by 10 additional epochs for fine-tuning. The loss function combined binary cross-entropy and Jaccard loss, with the Dice coefficient serving as the primary performance metric. Preprocessing included clipping CT values to the range [− 1024, 280] HU and normalizing them to [0, 1] using linear scaling. Data augmentation on training images, implemented using the Albumentations library (https://github.com/albumentations-team/albumentations), involved random rotations (0–360°, probability = 0.9), random-sized crops between 75 and 100% of the original area, horizontal flips (probability = 0.5), and nearest-neighbor interpolation to preserve masks. Slices were partitioned into training and validation sets using an 85/15 split, stratifying by the presence of high attenuation areas within the slice. Random shuffling was applied before splitting to ensure robust stratification. The implementation was carried out using TensorFlow/Keras and the segmentation_models library (https://github.com/qubvel/segmentation_models), with model checkpointing based on the highest validation Dice coefficient.

A patient-specific thresholding method is employed to refine the segmentation. First, applying a Gaussian filter enhances the local brightness distribution within the lungs. Then, thresholding is applied based on the modal density value, which is assumed to represent well-aerated tissue best, increased by 40 Hounsfield Units (HU).

The resulting pathological regions with high attenuation are then classified into distinct pathological categories, namely opacifications and reticulations/consolidations, using the conditional adaptive thresholding approach [30].

A morphological reconstruction combined with a pair of thresholds for seeding and masking is used to segment pathological areas characterized by decreased attenuation.

The vessel segmentation process, fully developed in MATLAB, is summarized in Fig. 3 and in Supplementary Fig. 5.

Preliminary segmentation is based on the Jerman filter [29]. To eliminate false positives, several criteria are applied, including volume thresholding, morphological reconstruction by seeding the inner lung region adjacent

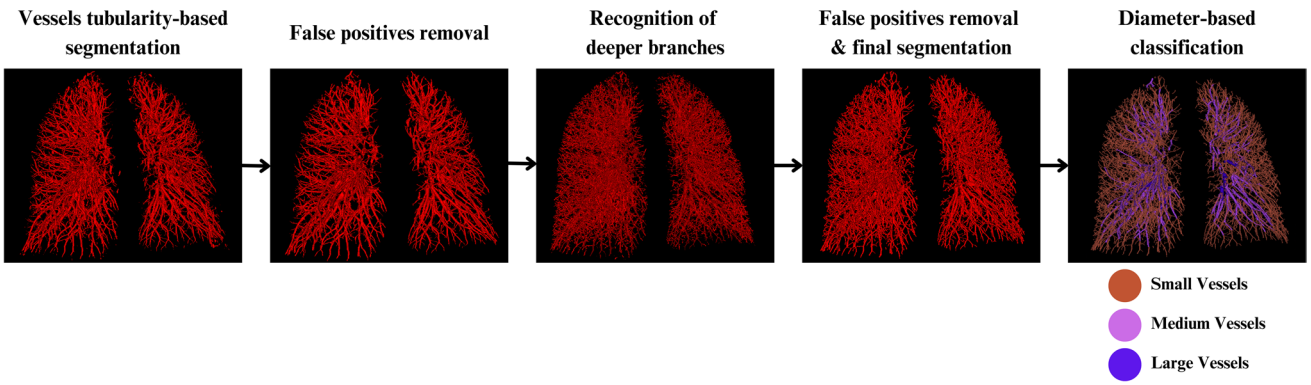


Fig. 3 Schematic representation of the vasculature segmentation algorithm

to the mediastinum, and additional constraints tailored to remove segmentation errors in critical areas—such as lung fissures and distal lung parenchyma—based on parameters like diameter, length (to account for regular branching enlargements), volume, anisotropy, and segregation. Moreover, a graph-based approach is implemented to discard convoluted segments in small branches that are unlikely to represent vascular structures, but potential leakages related to fibrotic bands.

To ensure comprehensive vessel segmentation, deeper branches are integrated by applying supplementary intensity thresholding along with additional Jerman filtering. Finally, the vascular segmentation is classified according to vessel diameter.

The pipeline ultimately measures the volumes of the segmented structures globally and separately for each lung and subregion. Beyond structural quantifications, the pipeline computes the gas volume and a coarse index of the air-to-perfusion ratio, calculated by dividing the gas volume by the vasculature volume.

Additional methodological details are available in the Supplementary material.

The basecode for the HRCT processing pipeline will be released on GitHub under an open-source license once it is ready for distribution.

Pipeline validation

The airway segmentation performance was tested on 8 cases from the public EXACT'09 dataset, while the pathological lung parenchyma segmentation was evaluated on the Radiopaedia and Coronacases datasets.

The overall pipeline was tested on 20 chest HRCT scans, encompassing both mild and severe cases of COVID-19 and IPF. Three radiologists (one with more than 10 years of experience and two fourth-year radiology residents) independently assessed the scans, blind to patient details. Each rater qualitatively scored the completeness, accuracy, and classification (GGO vs. reticulations/consolidations) of each segmentation target (airways, vessels, and parenchymal injury patterns) using a 5-point scale.

The segmentation results obtained by the proposed pipeline were also visually compared with those obtained using three available tools, namely LungCTAnalyzer, PTK, and TotalSegmentator.

Statistical Analysis

The mode [minimum–maximum] of the individual scores assigned by each of the three raters was recorded to obtain an overall assessment of the proposed methodology. The mean of the scores was also calculated to better describe potential differences in their ratings.

For each item, Kruskal–Wallis tests—accounting for data non-normality—were performed to assess potential differences in scores across raters, also stratifying by condition (COVID-19 vs. IPF) and severity (mild vs. severe). Statistical significance was set a p value of 0.05. To account for multiple comparisons, the Bonferroni correction was applied to adjust the significance threshold. Additionally, both strict and lenient percentages of agreement were reported to assess inter-rater reliability. The lenient agreement was defined as ratings within ± 1 point of each other, while strict agreement required exact matching scores.

All statistical analyses were performed using Python's statsmodels package (statsmodels v0.14.2).

Results

Results are presented in terms of qualitative evaluation of the segmentations, then quantitative evaluation of the individual structures, and visual rating of the overall segmentation by radiologists.

Qualitative evaluation of the segmentations. The proposed pipeline exhibits high performance in reconstructing

lung structures, such as airways and vasculature, and identifying patterns of injury in patients with infectious and inflammatory pathologies at different disease stages. Visual inspection demonstrates higher segmentation performance of the proposed pipeline compared to available tools typically employed in clinical studies. The outperformance is particularly evident when considering the proposed methodology's ability to distinguish between structures with similar imaging features, which are often confused by traditional segmentation approaches. The comparison between the proposed technique and three different alternatives, namely LungCTAnalyzer, PTK, and TotalSegmentator, is illustrated in Figs. 4, 5. Figure 4 compares two representative COVID-19 patients—one with mild and one with severe picture. Figure 5 provides a similar comparison for two representative IPF patients. Visual comparison reveals a superior reconstruction and differentiation performance in the segmentation provided by the proposed pipeline. Figure 6 visually compares the airway segmentation obtained using the proposed technique and the open-source U-Net model BronchiNet, implemented in the pipeline to deliver the preliminary segmentation. As shown, the proposed pipeline successfully recovers peripheral bronchial branches that are missed by BronchiNet alone, particularly in pathological case.

Quantitative Evaluation of Individual Structures on Different Datasets. Separate datasets were considered to evaluate the pipeline's performance on airways and parenchymal injury patterns. The pipeline was tested on 8 scans from the EXACT'09 dataset for airway segmentation, achieving a median Dice coefficient of 0.91 and a median IoU of 0.83 (Supplementary Table 2). The U-Net-based segmentation of the pathological regions with increased attenuation was evaluated on the Radiopaedia and Coronacases datasets, achieving a Dice coefficient of 0.90 on the training set and 0.89 on the test set.

Results of the Entire Pipeline on Multiple Structures as Rated by Radiologists. Table 1 summarizes ratings assigned by the three radiologists to the completeness and accuracy of airway and vessel segmentation and to the correctness of the classification of parenchymal injury patterns (GGO and reticulations/consolidations) on chest HRCT from 10 patients with COVID-19 (5 males; median age 56 [IQR 50–60] years) and 10 patients with IPF (10 males; median age 76 [IQR 73–77] years).

The airways achieved a segmentation completeness modal rate of 4 [4-5] and a higher segmentation accuracy of 5 [5-5]. Vessels achieved a segmentation completeness and accuracy rate of 4 [4-4] and 5 [5-5], respectively. Parenchymal alteration patterns achieved a completeness and accuracy rate of 4 [4-5] and 4 [4-4], respectively. The classification performance received a rating of 4 [4-4].

The results of the Kruskal–Wallis tests, which examined potential differences in scores across raters, diseases, and levels of severity, are presented in Table 2. Before correction for multiple comparisons, differences ($p < 0.05$) were observed in airways' completeness when considering the overall dataset ($p = 0.009$; strict agreement 56.7%, lenient agreement 100%), and specifically in the COVID-19 ($p = 0.016$; strict agreement 53.3%, lenient agreement 100%) and mild severity subgroups ($p = 0.006$; strict agreement 53.3%, lenient agreement 100%). However, after applying the Bonferroni correction, none of these differences remained statistically significant.

Analysis of inter-rater reliability revealed variable agreement patterns when considering a strict percentage of agreement, but a consistently high consensus using the lenient agreement measure. Airways' accuracy demonstrated the highest strict agreement (96.7% overall) with perfect lenient agreement (100%). For vasculature segmentation, completeness showed moderate strict agreement (75.0%) with near-perfect lenient agreement (98.3%), while accuracy demonstrated strict agreement of 73.3% and perfect lenient agreement (100%).

Concerning injury patterns, completeness exhibited partial strict agreement (53.3% and 33.3% in mild cases) and high lenient agreement (96.7% and 93.3%). Accuracy demonstrated strong strict agreement (83.3%) and high lenient agreement (96.7%). Classification had a strict agreement of 60.0%, with perfect lenient agreement (100%).

Notably, lenient agreement rates exceeded 93% for all items, indicating that raters predominantly provided ratings within one point of each other, which demonstrates practical consistency in the evaluation of the proposed methodology despite minor variations in exact scoring.

Supplementary Table 3 reports volumes and densities of the injury patterns segmented by the pipeline on the 20 chest HRCT from COVID-19 and IPF patients, offering additional insights into pathological tissue segmentation results.

Discussion

The proposed chest CT processing pipeline offers a novel and comprehensive approach for quantitatively investigating lung structures. Our method integrates traditional and AI-based segmentation techniques, leveraging the flexibility of AI for preliminary segmentations and the precision of traditional methods for fine-tuning results based on contextual knowledge. This approach has proven effective in providing accurate, robust, and potentially generalizable segmentation of pulmonary structures in patients with representative inflammatory and infectious diseases, namely COVID-19 and IPF, at different severity levels.

Qualitative evaluation of the segmentations obtained by the proposed pipeline shows high performance in reconstructing lung structures, including airways and

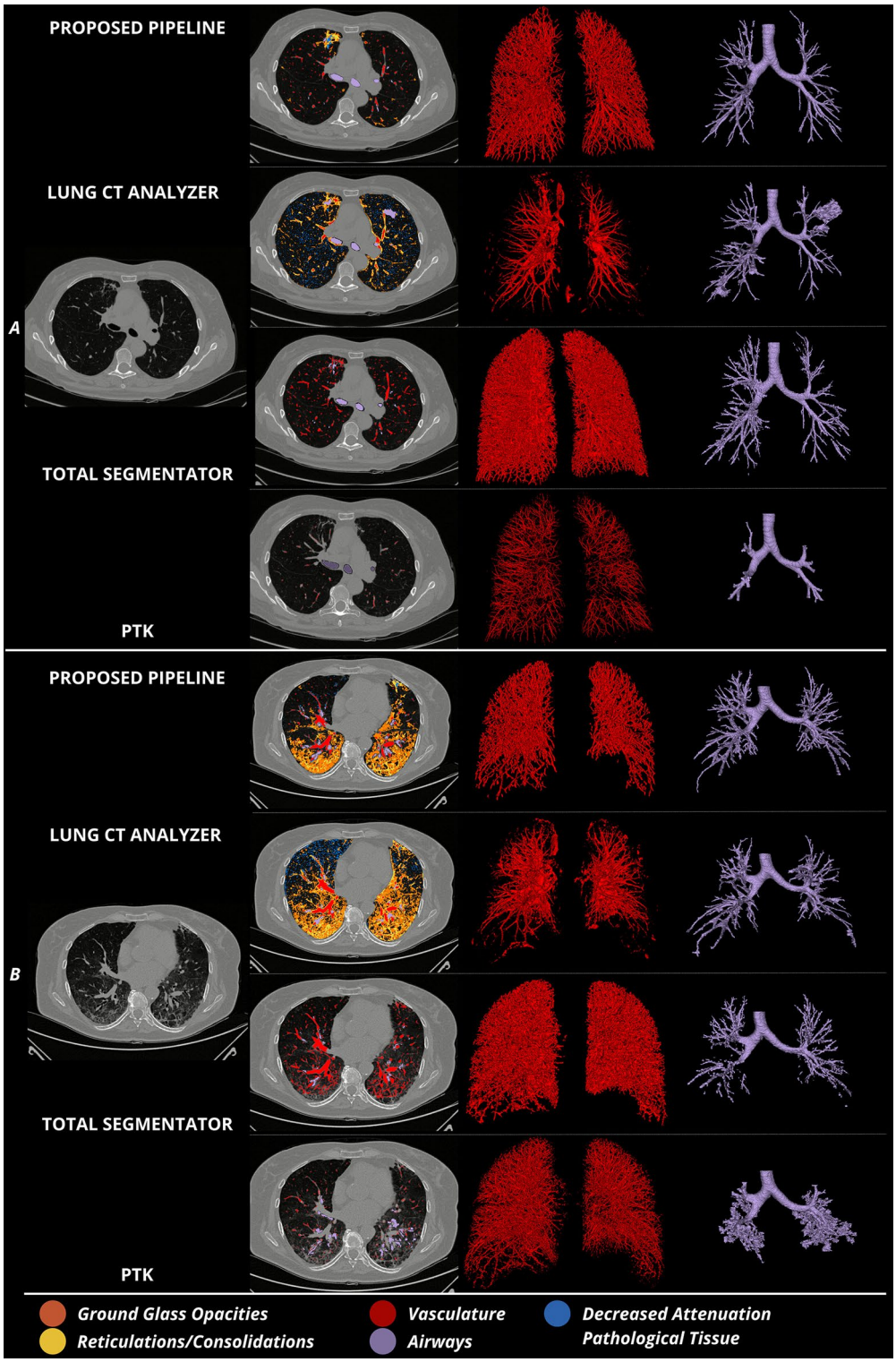


Fig. 4 Comparison of segmentations generated by the proposed technique and three alternative available methods—LungCTAnalyzer, TotalSegmentator, and PTK—on two representative COVID-19 patients, with either mild (A: 55-year-old female) or severe pathology (B: 67-year-old female)

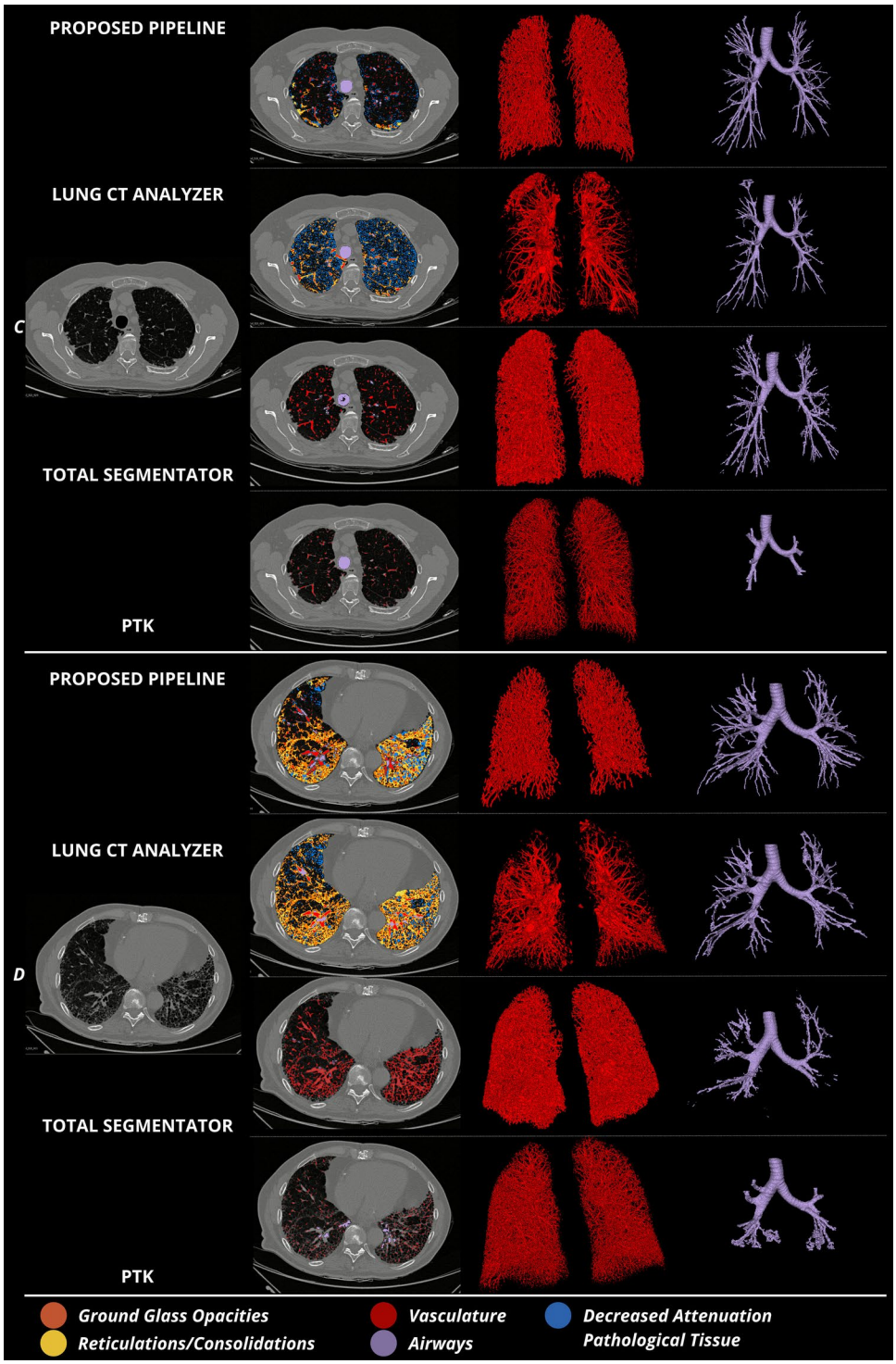


Fig. 5 Comparison of segmentations generated by the proposed technique and three alternative available methods—LungCTAnalyzer, TotalSegmentator, and PTK—on two representative idiopathic pulmonary fibrosis patients, with either mild (A: 77-year-old male) or severe pathology (B: 72-year-old male)

vasculature, and identifying patterns of injuries in patients with infectious and inflammatory pathologies at different disease stages. Visual inspection (Figs. 4 and 5) indicates superior segmentation performance of the proposed pipeline compared to tools often used in clinical studies. This outperformance is primarily marked in the methodology's capability to differentiate between structures with similar imaging features, frequently confused by traditional segmentation approaches.

The multi-purpose approach allows for leveraging insights gathered from each individual task to facilitate the correct identification of different structures in the lung that

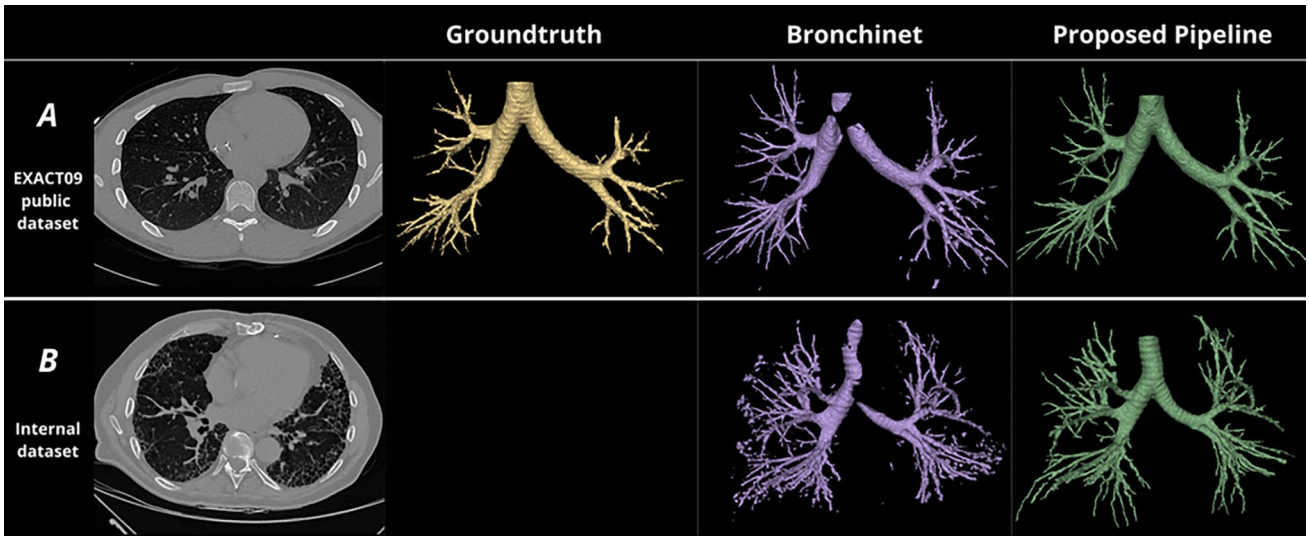


Fig. 6 Comparison of airway segmentation obtained using the proposed technique and the open-source U-Net model BronchiNet. Figure A of the panel shows the segmentation performed on a CT scan from the publicly available EXACT'09 dataset (Exact 01), for which the ground truth mask is provided. Figure B shows the segmentation obtained from a CT with severe honeycombing belonging to the internal validation dataset (72-year-old male). This second figure highlights the ability of the proposed pipeline to integrate deep airways even in extensively pathological pictures

might be missed or mistaken by single-purpose strategies focusing on a unique type of pulmonary structure.

The high-quality segmentations were validated through quantitative assessment of individual pulmonary structures (airways and injury patterns) using annotated datasets. Quantitative validation of vessel segmentation was not performed due to the lack of annotated datasets available for the analysis. The only available dataset, to our knowledge, is VESSEL12 (https://vessel12.grand-challenge.org/) [31], which provides only some classified coordinates also located in very deep vascular branches.

Evaluation of the overall pipeline, performed by three independent radiologists using a 5-point scale, showed high performance across all evaluated items and criteria, including the completeness and accuracy of the airways, pulmonary vessels, and parenchyma injury, as well as the correct classification of GGO and reticulations/consolidations.

All segmentations received scores of 3 or higher, with mode values of 4 for all items. The accuracy of parenchyma injury patterns had the lowest mean score of 3.8, possibly due to the occasional overestimation of pathological regions at vascular terminations below the resolution limit. Notably, airway segmentation accuracy received the highest scores.

Statistical analysis revealed no significant differences in ratings across raters, diseases, or severity levels after applying Bonferroni correction for multiple comparisons, suggesting consistent performance of the methodology across different clinical scenarios.

Inter-rater reliability, assessed through strict and lenient percentage of agreement, demonstrated variable patterns across items. Strict agreement for the overall dataset ranged from 53.3% to 96.7%, while lenient agreement (allowing ratings within one point of each other) consistently exceeded 93%, reaching 100% for several categories. The lowest value of strict agreement concerns the completeness of injury patterns in mild cases, which may be due to the fact that in such cases the lesion patterns are often small, mild, and scattered. Raters may disagree on whether the automated segmentation captured all subtle opacities or missed some marginal findings. Moreover, the transition from altered to healthy tissue is more gradual, leading to difficult interpretation of whether the segmentation extent is "complete" or if it under- or overestimated the true extent.

Nonetheless, the highly lenient agreement indicates that raters predominantly provided similar evaluations, with differences typically limited to adjacent rating points.

While acknowledging that percentages of agreement do not account for chance, more advanced metrics, such as Cohen's Kappa, cannot be considered fully reliable given our study design, which involves only three raters and items that exhibit restricted variability.

These findings support the technique's potential in studying various infectious and inflammatory pathological conditions.

The main limitation of the developed technique is the extensive number of rules, criteria, and parameters, which have been selected empirically. The trial-and-error approach has been applied across various cases to identify an optimal configuration. Future efforts are needed to make the optimization process more systematic, assess the parameters'

Table 1 Evaluation of the segmentation performance of the proposed pipeline across raters and pathological conditions

Validation item	Rater	Whole group ($n=20$)	COVID-19 ($n=10$)	IPF ($n=10$)
Airways completeness	Rater 1	4[3–5]	4[4–5]	4[3–5]
	Rater 2	5[3–5]	5[4–5]	4[3–5]
	Rater 3	4[4–4]	4[4–4]	4[4–4]
	Overall	**4[4–5]**	**4[4–5]**	**4[4–4]**
Airways accuracy	Rater 1	5[5–5]	5[5–5]	5[5–5]
	Rater 2	5[4–5]	5[5–5]	5[4–5]
	Rater 3	5[5–5]	5[5–5]	5[5–5]
	Overall	**5[5–5]**	**5[5–5]**	**5 [5–5]**
Vessels completeness	Rater 1	4 [4–5]	4 [4–5]	4 [4–5]
	Rater 2	4 [4–5]	5 [4–5]	4 [4–5]
	Rater 3	4[3–5]	4 [4–5]	4 [3–5]
	Overall	**4 [4–4]**	**4[4–5]**	**4 [4–4]**
Vessels accuracy	Rater 1	5 [4–5]	5 [4–5]	−[4–5]
	Rater 2	5 [4–5]	5 [4–5]	−[4–5]
	Rater 3	5 [4–5]	5 [4–5]	−[4–5]
	Overall	**5 [5–5]**	**5 [5–5]**	−[4–5]
Injury patterns completeness	Rater 1	5 [3–5]	5 [4–5]	4 [3–5]
	Rater 2	4 [3–5]	4 [4-5]	− [3–5]
	Rater 3	4 [3–5]	− [3–5]	4 [3–5]
	Overall	**4 [4–5]**	**− [4–5]**	**4 [4–4]**
Injury patterns accuracy	Rater 1	4 [3–4]	4 [3–4]	4 [3–4]
	Rater 2	4 [3–5]	4 [3–5]	4 [3–4]
	Rater 3	4 [3–4]	− [4–5]	4 [3–4]
	Overall	**4 [4–5]**	**4 [4–4]**	**4 [4–4]**
Injury patterns classification	Rater 1	4 [3–5]	4 [4–5]	4 [3–5]
	Rater 2	− [3–5]	5 [3–5]	4 [3–5]
	Rater 3	4 [3–5]	4 [3–5]	4 [3–4]
	Overall	**4 [4–5]**	**4 [4–5]**	**4 [4–4]**

The table includes ratings of airways, vessels, and alteration patterns' (GGO and reticulations/consolidations) segmentation assigned by three raters using a 5-point scale (with 1 = lowest and 5 = highest score), as well as overall ratings, aggregating modal scores across raters. The evaluation was performed on chest HRCT scans from 20 patients with either COVID-19 ($n=10$) or idiopathic pulmonary fibrosis (IPF, $n=10$) at different disease severity. Scores are reported as mode [minimum—maximum]. In case of no mode, the sign "-" was reported

impact through a sensitivity analysis, and validate the segmentation pipeline on a broader range of pulmonary pathologies and severity levels.

Additionally, the proposed segmentation method, followed by manual refinement where necessary, could facilitate the creation of an extended dataset with precise annotations and aid the development of a fully AI-based multi-class segmentation model capable of handling pathological cases, thus overcoming the flaws of currently available public AI models.

It is important to note that our pipeline was developed and validated specifically for diffuse infectious and inflammatory lung diseases. Its application to non-inflammatory or obstructive conditions, such as lung tumors or emphysema, is not supported by our current findings and would likely require architectural modifications or recalibrations on the pipeline and models retraining on relevant datasets. Future work will explore the transferability of our approach's core components to these other important disease categories.

Further research directions could explore the integration of advanced 3D Transformer-based architectures and semi-supervised learning strategies, which have demonstrated remarkable success in segmenting specific pulmonary and pathological structures [15, 18–20, 23, 25].

In conclusion, the chest CT processing pipeline we developed stands as a comprehensive and flexible solution for quantitatively characterizing possible alterations in lung architecture. It provides valuable information to complement clinical and laboratory data. Indeed, detailed insights into lung architecture can enhance our

Table 2 Stratified inter-rater reliability analysis across raters, diseases, and severity levels

Items	Effects	*H*	*p*	*p* corrected	Percentage of agreement	Percentage of agreement (± 1 score deviation)
Airways completeness	All cases	9.347	0.009	0.327	56.7	100.0
	COVID-19	8.286	0.016	0.556	53.3	100.0
	IPF	2.542	0.281	1.000	60.0	100.0
	Mild	10.268	0.006	0.206	53.3	100.0
	Severe	1.295	0.523	1.000	60.0	100.0
Airways accuracy	All cases	2.000	0.368	1.000	96.7	100.0
	COVID-19	0.000	1.000	1.000	100.0	100.0
	IPF	2.000	0.368	1.000	93.3	100.0
	Mild	0.000	1.000	1.000	100.0	100.0
	Severe	2.000	0.368	1.000	93.3	100.0
Vessels completeness	All cases	0.733	0.693	1.000	75.0	98.3
	COVID-19	1.837	0.399	1.000	80.0	100.0
	IPF	1.725	0.422	1.000	70.0	96.7
	Mild	0.806	0.669	1.000	66.7	100.0
	Severe	0.450	0.799	1.000	83.3	96.7
Vessels accuracy	All cases	0.000	1.000	1.000	73.3	100.0
	COVID-19	0.000	1.000	1.000	73.3	100.0
	IPF	0.000	1.000	1.000	73.3	100.0
	Mild	1.208	0.547	1.000	73.3	100.0
	Severe	0.773	0.679	1.000	73.3	100.0
Injury patterns completeness	All cases	2.054	0.358	1.000	53.3	96.7
	COVID-19	1.652	0.438	1.000	56.7	96.7
	IPF	0.548	0.760	1.000	50.0	96.7
	Mild	3.654	0.161	1.000	33.3	93.3
	Severe	0.039	0.981	1.000	73.3	100.0
Injury patterns accuracy	All cases	1.908	0.385	1.000	83.3	96.7
	COVID-19	2.608	0.271	1.000	80.0	93.3
	IPF	0.000	1.000	1.000	86.7	100.0
	Mild	4.108	0.128	1.000	73.3	93.3
	Severe	1.036	0.596	1.000	93.3	100.0
Injury patterns classification	All cases	2.845	0.241	1.000	60.0	100.0
	COVID-19	1.608	0.448	1.000	66.7	100.0
	IPF	1.483	0.476	1.000	53.3	100.0
	Mild	3.299	0.192	1.000	53.3	100.0
	Severe	0.323	0.851	1.000	66.7	100.0

Results are presented for the overall dataset ("All cases") and stratified by disease subgroup (COVID-19 vs. IPF) and severity level (mild vs. severe). The nonparametric Kruskal–Wallis test was used to account for data non-normality, with statistical significance assessed using both uncorrected and multiple comparison-corrected *p* values obtained with the Bonferroni method, which accounted for the 35 comparisons. The strict percentage of agreement and a lenient version, allowing ± 1 score deviation, are reported. The significance level was set to $p = 0.05$

understanding of lung disease pathophysiology, foster accurate diagnosis, prognosis, and monitoring, and improve therapeutic decision-making across various lung pathologies. Furthermore, our tool holds promise for radiomics studies. Providing accurate and robust lung structure segmentation lays the groundwork for the high-throughput extraction of image features, which is at the core of radiomics. This could further enhance the diagnostic and prognostic potential of chest HRCT in pulmonary diseases, thereby contributing to personalized medicine.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11547-025-02166-w.

Acknowledgements A. Arrigoni would like to express his gratitude to the "Aiuti per la Ricerca sulle Malattie Rare" (A.R.M.R) Foundation.

Funding This study was supported in part by a grant from Brembo S.p.A. (Curno, Bergamo, Italy) under the initiative "Progetto TrexUno." A. Arrigoni was supported by a scholarship from the "Aiuti per la Ricerca sulle Malattie Rare" (A.R.M.R.) Foundation.

Data availability The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval The study was conducted in accordance with the Declaration of Helsinki and within the framework of the observational study protocols Reg. No. 2023/0233, and Reg. No. 2023/0047, approved by the local Institutional Review Board (Comitato Etico di Bergamo, Italy). All subjects provided written informed consent at clinical study enrollment.

References

1. Elicker BM, Kallianos KG, Henry TS (2017) The role of high-resolution computed tomography in the follow-up of diffuse lung disease. Eur Respir Rev 26:170008. https://doi.org/10.1183/16000617.0008-2017
2. Müller NL (2002) Computed tomography and magnetic resonance imaging: past, present and future. Eur Respir J Suppl 35:3s–12s
3. Balbi M, Conti C, Imeri G et al (2021) Post-discharge chest CT findings and pulmonary function tests in severe COVID-19 patients. Eur J Radiol 138:109676. https://doi.org/10.1016/j.ejrad.2021.109676
4. Corsi A, Caroli A, Bonaffini PA et al (2022) Structural and functional pulmonary assessment in severe COVID-19 survivors at 12 months after discharge. Tomography 8:2588–2603. https://doi.org/10.3390/tomography8050216
5. Adegunsoye A, Oldham JM, Bellam SK et al (2019) Computed tomography honeycombing identifies a progressive fibrotic phenotype with increased mortality across diverse interstitial lung diseases. Ann Am Thorac Soc 16:580–588. https://doi.org/10.1513/AnnalsATS.201807-443OC
6. Hu Q, Liu Y, Chen C et al (2021) Reversible bronchiectasis in COVID-19 survivors with acute respiratory distress syndrome: pseudobronchiectasis. Front Med. https://doi.org/10.3389/fmed.2021.739857
7. Primack SL, Müller NL, Mayo JR et al (1994) Pulmonary parenchymal abnormalities of vascular origin: high-resolution CT findings. Radiographics. https://doi.org/10.1148/radiographics.14.4.7938765
8. Imeri G, Conti C, Caroli A et al (2024) Gas exchange abnormalities in Long COVID are driven by the alteration of the vascular component. Multidisciplinary Respiratory Med. https://doi.org/10.5826/mrm.2024.938
9. Walsh SLF, Hansell DM (2014) High-Resolution CT of interstitial lung disease: a continuous evolution. Semin Respir Crit Care Med 35:129–144. https://doi.org/10.1055/s-0033-1363458
10 Cabral RH, Branco É, dos Rizzo MS et al (2011) Cell therapy for fibrotic interstitial pulmonary disease: experimental study. Microscopy Res Tech. https://doi.org/10.1002/jemt.20981
11. Sverzellati N (2013) Highlights of HRCT imaging in IPF. Respir Res 14(Suppl 1):S3. https://doi.org/10.1186/1465-9921-14-S1-S3
12. Li H, Tang Z, Nan Y, Yang G (2022) Human treelike tubular structure segmentation: a comprehensive review and future perspectives. Comput Biol Med 151:106241. https://doi.org/10.1016/j.compbiomed.2022.106241
13. Garcia-Uceda A, Selvan R, Saghir Z et al (2021) Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks. Sci Rep 11:16001. https://doi.org/10.1038/s41598-021-95364-1
14. Hofmanninger J, Prayer F, Pan J et al (2020) Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp 4:50. https://doi.org/10.1186/s41747-020-00173-2
15. Wu R, Xin Y, Qian J, Dong Y (2023) A multi-scale interactive U-Net for pulmonary vessel segmentation method based on transfer learning. Biomed Signal Process Control 80:104407. https://doi.org/10.1016/j.bspc.2022.104407
16. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation
17. Isensee F, Jaeger PF, Kohl SAA et al (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18:203–211. https://doi.org/10.1038/s41592-020-01008-z
18. Yadav DP, Sharma B, Webber JL et al (2024) EDTNet: a spatial aware attention-based transformer for the pulmonary nodule segmentation. PLoS ONE 19:e0311080. https://doi.org/10.1371/journal.pone.0311080
19. Liu H, Zhuang Y, Song E et al (2024) A 3D boundary-guided hybrid network with convolutions and transformers for lung tumor segmentation in CT images. Comput Biol Med 180:109009. https://doi.org/10.1016/j.compbiomed.2024.109009
20. Liu Z, Li W, Cui Y et al (2025) Label-efficient transformer-based framework with self-supervised strategies for heterogeneous lung tumor segmentation. Expert Syst Appl 269:126364. https://doi.org/10.1016/j.eswa.2024.126364
21. Wang D, Han C, Zhang Z et al (2024) FedDUS: lung tumor segmentation on CT images through federated semi-supervised with dynamic update strategy. Comput Methods Programs Biomed 249:108141. https://doi.org/10.1016/j.cmpb.2024.108141
22. Qiao P, Li H, Song G et al (2023) Semi-supervised CT lesion segmentation using uncertainty-based data pairing and SwapMix. IEEE Trans Med Imaging 42:1546–1562. https://doi.org/10.1109/TMI.2022.3232572
23. Sharafeldeen A, Khelifi A, Ghazal M et al (2025) Unsupervised Segmentation of Pulmonary Regions in 3D CT Scans Optimized Using Transformer Model. In: Antonacopoulos A, Chaudhuri S, Chellappa R et al (eds) Pattern Recognition. Springer Nature Switzerland, Cham, pp 51–66
24. Chen L, Tang R, Chen H et al (2021) Pulmonary vasculature: a target for COVID-19. Am J Respir Crit Care Med 203:260–261. https://doi.org/10.1164/rccm.202009-3564LE

25. Yang J, Li H, Wang H, Han M (2024) 3D medical image segmentation based on semi-supervised learning using deep co-training. Appl Soft Comput 159:111641. https://doi.org/10.1016/j.asoc.2024.111641
26. Qin Y, Zheng H, Gu Y et al (2021) Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT. IEEE Trans Med Imaging 40:1603–1617. https://doi.org/10.1109/TMI.2021.3062280
27. Zhang M, Wu Y, Zhang H et al (2023) Multi-site, Multi-domain Airway Tree Modeling. Med Image Anal 90:102957. https://doi.org/10.1016/j.media.2023.102957
28. Vincent L (1993) Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans Image Process 2:176–201. https://doi.org/10.1109/83.217222
29. Jerman T, Pernuš F, Likar B, Špiclin Ž (2016) Enhancement of vascular structures in 3D and 2D angiographic images. IEEE Trans Med Imaging 35:2107–2118. https://doi.org/10.1109/TMI.2016.2550102
30. Horváth A, Spindler S, Szalay M, Rácz I (2016) Preprocessing endoscopic images of colorectal polyps. Acta Tech Jaurinensis 9:65–82. https://doi.org/10.14513/actatechjaur.v9.n1.397
31. Rudyanto RD, Kerkstra S, van Rikxoort EM et al (2014) Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study. Med Image Anal 18:1217–1232. https://doi.org/10.1016/j.media.2014.07.003