npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-026-02683-6

Deep-learning based quantitative evaluation of postoperative atelectasis following right upper lobectomy

Devanish N. Kamtam[1] ✉, Giuseppe M. Facchi[2,3], Nicole Lin[1], Lillian L. Tsai[1], Natalie S. Lui[1], Irmina A. Elliott[1,4], Doug Z. Liou[1], Leah M. Backhus[1,4], Mark F. Berry[1], Henry H. Guo[5], Curtis P. Langlotz[5,6,7] & Joseph B. Shrager[1,4]

Existing methods of grading atelectasis are typically subjective and not scalable. We aimed to develop an automated, deep learning–based framework to quantify and grade postoperative atelectasis. We retrospectively included all patients who underwent RULobectomy from 2008 to 2023. We trained three nnU-Net v2 segmentation models for preoperative and postoperative lobes and airways with volumetric quantification of the right middle lobe (RML), right lower lobe (RLL), and total lung volume. Atelectasis severity in the RML was independently graded using a 5-point radiological scale (none, minimal, subsegmental, segmental, lobar). The association between volume metrics with atelectasis severity and clinical outcomes was evaluated. 236 patients comprised the study cohort. Median(IQR) RML volume loss progressively increased with higher atelectasis grades, from –4.6 mL (−78.5, 59.0) in grade 0 to –317.8 mL (−440.7, −194.8) in grade 4 atelectasis ($p < 0.001$). Normalized RML/right lung (RL) and RML/total lung (TL) volume ratios showed statistically significant differences across the pooled atelectasis grades ($p < 0.001$). Normalized RLL volumes increased with worsening RML atelectasis ($p < 0.001$), suggesting compensatory hyperinflation. A higher ΔRML/RL [OR(95%CI): 0.89 (0.81–0.98), $p = 0.01$] and ΔRML/TL [0.80 (0.65–0.98), $p = 0.03$] were associated with reduced 1-year need for bronchoscopy. We demonstrate the feasibility and clinical relevance of deep learning–based volumetric assessment of atelectasis after RULobectomy.

Atelectasis—partial or complete collapse of a lung lobe resulting from alveolar deflation or inadequate expansion—is one of the most common pulmonary complications following surgery[1]. It is particularly common following cardiothoracic surgery, with rates of 30–72%[2,3], versus 3.2% following non-cardiothoracic surgery[4]. Atelectasis can be classified into obstructive (resorptive), compressive, and adhesive types, based on underlying mechanisms such as internal airway obstruction, external compression, or surfactant dysfunction. Collapse of lung lobes reduces the alveolar surface area available for gas exchange, leading to impaired oxygenation, hypoxemia, and, in severe cases, respiratory failure. Further, completely atelectatic lung lobes and segments have a propensity to become infected.

While early postoperative atelectasis is often self-limited, complete lobar collapse and ongoing, substantial sublobar collapse can lead to significant morbidity, especially following lung resections, where the remaining lung must compensate for the loss of resected parenchyma. Postoperative atelectasis is typically identified through radiographic imaging, with chest X-rays for preliminary assessment, followed by CT scans for diagnostic confirmation and more detailed assessment of the severity of atelectasis.

To facilitate consistent evaluation of atelectasis—albeit largely in non-postoperative scenarios—several studies have proposed standardized grading criteria[5], some of which have also been validated against clinical outcomes[6,7]. Early objective approaches to grading atelectasis primarily

[1]Division of Thoracic Surgery, Department of Cardiothoracic Surgery, Stanford University School of Medicine, Stanford, CA, USA. [2]Division of Head and Neck Surgery, Department of Otolaryngology, Stanford University, Stanford, CA, USA. [3]Department of Computer Science, University of Milan, Milan, Italy. [4]Department of Surgery, Veterans Affairs Palo Alto Health Care System, Stanford, CA, USA. [5]Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA. [6]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [7]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA. ✉e-mail: devanish@stanford.edu

relied on Hounsfield Unit (HU) measurements[8–11]. CT slice intensity values between −500 and +100 HU (vs. <−500 HU as aerated lung) have been used to identify poorly aerated regions reflecting atelectasis[8–11], and these metrics demonstrated good correlation with clinical outcomes. While effective within the lung parenchyma, this approach becomes less reliable when atelectasis extends to the lung periphery, as the collapsed lung border may be difficult to distinguish from adjacent mediastinal structures on CT. And in complete lobar collapse, the proportion of collapsed voxels may be outweighed by the volume of the remaining aerated lung, leading to underestimation of atelectasis severity. Moreover, despite HU standardization (−1000 HU—air, 0 HU—water), the HU thresholds may differ by scanner type/imaging acquisition protocols (Supplementary Fig. 1).

More recently, other methods, such as the BEST-CT[12], have been used to quantify atelectasis as a percentage of total lung volume. This method divides axial CT slices into grids and classifies each segment into one of ten status categories, including atelectasis, the percentage of which is then calculated across the entire scan. While this methodology has been validated in two clinical contexts—cystic fibrosis[13] and bronchiectasis[12]—the grading remains entirely manual and time-intensive. Recently, less time-intensive scoring systems like the ASSESS criteria have been introduced in post-interventional settings, such as following bronchoscopy under general anesthesia, grading atelectasis by its extent on CT from the posterior chest wall to the anterior vertebral border. Their applicability, however, is limited to dependent dorso-caudal lung regions.

Overall, existing methods of grading atelectasis remain constrained by their manual, condition-specific, subjective nature, with significant inter-rater variability. This limits their generalizability, scalability, and reproducibility. Deep learning may enable more robust and automated grading of atelectasis that is reproducible across diverse clinical contexts, imaging protocols, and healthcare systems—provided such variability is adequately represented in the training data. Atelectasis can be graded either by segmenting and quantifying the reduction in volume of the collapsed regions of the lung[5,8] or, alternatively, by segmenting and quantifying volume of the aerated regions that represent the remaining functional lung. Prior studies[14,15] have confirmed that quantitative volumetric analysis of lung aeration—whether by CT or MRI[15,16]—provides objective metrics that can serve as reliable proxies for assessing the degree of atelectasis. Hence, we aimed to use automated volumetric analysis of aerated lung regions to quantify and grade the severity of atelectasis.

Given its higher incidence following thoracic surgery, we evaluated atelectasis following pulmonary lobectomy, using right upper lobectomy (RULobectomy)—the most frequently performed lobectomy—as a representative model. This was also an ideal model because chronic right middle lobe (RML) atelectasis, often referred to as right middle lobe syndrome, is a recognized long-term complication after right upper lobectomy[17]. Given this known risk, we specifically targeted the quantification of middle lobe volume loss and its correlation with physician-assigned atelectasis grades.

We leveraged paired preoperative and six-month-postoperative CT scans from patients who had undergone RULobectomy to develop and validate an automated, deep learning–based volumetric pipeline to quantify/grade atelectasis. Our aim was to establish a reproducible and clinically interpretable framework for quantifying lobar volume changes and grading postoperative atelectasis and to evaluate its clinical relevance by comparing volumetry-derived atelectasis metrics with physician-assigned grades in a large postoperative cohort. This approach may offer a scalable tool for both retrospective research and prospective clinical risk stratification, with potential applicability to other lobectomies and clinical contexts in which atelectasis occurs.

Results

Baseline cohort characteristics

Baseline demographic and clinical characteristics are summarized in Table 1. The median age was 69.2 years (IQR: 61.0–74.6), and 62.3% of patients were female. Common comorbidities included diabetes mellitus (17.8%), chronic obstructive pulmonary disease (16.5%), cerebrovascular disease (11.4%), and chronic kidney disease (11.0%).

Table 1 | Baseline demographic and clinical characteristics of all patients who underwent right upper lobectomy and those who were included in the final study cohort

Characteristics		Right upper lobectomy cohort (*n* = 438)	Study cohort (*n* = 236)
Age (years) (Median, IQR)		69.7(62.7–75.0)	69.2(61.0–74.6)
Sex	Male	179(40.9%)	89(37.9%)
	Female	259(59.1%)	147(62.3%)
Race	White	282(64.4%)	144(61.0%)
	Black	7(1.6%)	5(2.1%)
	Asian	98(22.4%)	64(27.1%)
	Other	43(9.8%)	23(9.7%)
	Unknown	8(1.8%)	0(0%)
Approach	Open	117(26.7%)	50(21.2%)
	VATS	248(56.6%)	134(56.8%)
	Robot-assisted	73(16.7%)	52(22.0%)
FEV_1 (%) (Median, IQR)		95(79.2–108)	97(80–109)
Diabetes mellitus		64(14.6%)	42(17.8%)
Chronic obstructive pulmonary disease		70(16.0%)	39(16.5%)
Congestive heart failure		25(5.7%)	15(6.4%)
History of cerebrovascular accident		45(10.3%)	27(11.4%)
Chronic kidney disease		40(9.1%)	26(11.0%)
Charlson-comorbidity score (Median, IQR)		6(4-7)	6(5-8)
Operative diagnosis	Adenocarcinoma	329(75.1%)	191(80.9%)
	SCC	43(9.8%)	14(5.9%)
	NEC	16(3.7%)	6(2.5%)
	Mixed	10(2.3%)	3(1.3%)
	Metastases	17(3.9%)	11(4.7%)
	Granulomatous inflammation	7(1.6%)	3(1.3%)
	Other benign lesions	16(3.7%)	8(3.4%)
Median time from preop scan to surgery (days) (Median, IQR)		44.5(24.0–74.0)	
Median time from surgery to postop scan (days) (Median, IQR)		196(173.2–207.7)	
Atelectasis on CT scan at 6 months	0	260(78.3%)	186 (78.8%)
	1	48(14.5%)	32 (13.6%)
	2	13(3.9%)	10 (4.2%)
	3	6(1.8%)	6 (2.5%)
	4	5(1.5%)	2 (0.8%)

Radiological data characteristics

The median time between preoperative CT scans and the surgery was 44.5 (24, 74) days, and the median time between the surgery and postoperative CT scans was 196 (173.2, 207.7) days. The median slice thickness was 1.25 (1.0, 1.25) mm for preoperative and postoperative scans. RML atelectasis at 6 months was absent in 186 (78.8%) of patients, while grade 1, 2, 3, and 4 RML atelectasis were observed in 32 (13.6%), 10 (4.2%), 6 (2.5%), and 2 (0.8%) patients.

Preoperative model performance

On the training dataset ($n = 93$), Dice scores ranged from 0.97 ± 0.01 for the RML to 0.99 ± 0.00 for the RLL, with an overall mean Dice of 0.98 ± 0.01.

Table 2 | Preoperative lobe, postoperative lobe, and preoperative airway segmentation model performance on CT scans, reported as Dice scores for each lung lobe and overall mean Dice score

Preoperative Lung lobes	Training dataset (93)	Internal test dataset (29)		External validation dataset (55)	
		Model inference	After post-processing	Model inference	After post-processing
Right upper lobe	0.98 ± 0.01	0.97 ± 0.03	0.97 ± 0.03	0.94 ± 0.07	0.95 ± 0.06
Right middle lobe	0.97 ± 0.01	0.96 ± 0.03	0.96 ± 0.03	0.86 ± 0.22	0.86 ± 0.22
Right lower lobe	0.99 ± 0.01	0.98 ± 0.04	0.98 ± 0.03	0.93 ± 0.12	0.94 ± 0.12
Left upper lobe	0.99 ± 0.00	0.98 ± 0.02	0.98 ± 0.02	0.93 ± 0.13	0.95 ± 0.13
Left lower lobe	0.99 ± 0.01	0.98 ± 0.03	0.98 ± 0.01	0.89 ± 0.18	0.91 ± 0.18
Overall mean	0.98 ± 0.01	0.97 ± 0.02	**0.98 ± 0.02**	0.91 ± 0.16	**0.92 ± 0.15**
Postoperative Lung lobes	Training dataset (65)	Internal test dataset (17)		External validation dataset	
		Model inference	After post-processing	-	-
Right middle lobe	0.96 ± 0.1	0.98 ± 0.01	0.98 ± 0.01	-	-
Right lower lobe	0.98 ± 0.01	0.99 ± 0.00	0.99 ± 0.00	-	-
Left lung	0.99 ± 0.02	0.99 ± 0.00	0.99 ± 0.00	-	-
Overall mean	0.98 ± 0.04	0.99 ± 0.00	**0.99 ± 0.00**	-	-
Airways	Training dataset (34)	Internal test dataset (7)		External validation dataset (27)	
Preoperative	0.94 ± 0.02	0.95 ± 0.01		0.87 ± 0.03	
	Training dataset (59)	Internal test dataset (22)		External validation dataset	
Postoperative	0.95 ± 0.01	0.94 ± 0.02			

The internal test dataset ($n = 29$) showed similarly strong performance, achieving an overall mean Dice score of 0.98 ± 0.02 after postprocessing (Table 2). The RML consistently yielded the lowest Dice scores across training and test datasets, likely reflecting the higher likelihood of incomplete right horizontal fissure vs. other fissures and smaller volume of this lobe.

On the external test dataset of LOLA11 challenge ($n = 55$), baseline model performance showed moderate variability across lobes, with Dice scores ranging from 0.86 ± 0.22 for the RML to 0.94 ± 0.07 for the RUL. Postprocessing yielded minor improvements in the mean Dice score from 0.91 ± 0.16 to 0.923 ± 0.15. This score is currently the joint state-of-the-art in the most comprehensive and longest running lung lobe segmentation challenge (0.928 ± 0.15)[18].

Postoperative model performance

On the training dataset ($n = 65$), Dice scores ranged from 0.96 ± 0.1 for the RML to 0.99 ± 0.02 for the left lung, with an overall mean Dice score of 0.98 ± 0.04 (Table 2). Internal testing ($n = 17$) yielded excellent performance, with an overall Dice score of 0.99 ± 0.00 after postprocessing. Among the individual lobes, the RML showed slightly lower Dice scores.

Automated inference, volumetric analysis, and atelectasis grading in clinical CT dataset

We applied our preoperative and post-operative model inference to the entire clinical CT dataset ($n = 236$). Eight postoperative scans (3.4%) were excluded from analysis owing to pleural pathologies, including effusion and pneumothorax. Owing to the lack of a reference standard in the postoperative dataset, all output masks were visually inspected, and manual corrections were required in 16 scans (6.8%) to ensure accurate lobar volumetry for atelectasis assessment. The median Dice score between the edited and original masks ($n = 16$) was 0.85 (IQR, 0.74–0.94).

We then extracted lobar volumes and assessed longitudinal lobar volume change using various metrics (Table 3). Median (IQR) ΔRML volume change demonstrated a progressive decline with increasing RML atelectasis grades from −4.6 mL (−78.5, 59.0) (no atelectasis) to −317.8 mL (−440.7, −194.8) (grade 4 atelectasis) ($p < 0.001$). This pattern persisted across normalized metrics, with ΔRML/RL and ΔRML/TL volume changes showing increasingly negative values with increasing grades of atelectasis ($p < 0.001$). Importantly, RLL volume changes showed an opposite trend, highlighting the expected compensatory hyperinflation of the RLL with increasing degrees of RML atelectasis. ΔRLL/RL volumes increased with worsening RML atelectasis (e.g., median ΔRLL/RL : 31.2% to 50.5%, $p < 0.001$) (Table 3). However, the ΔRLL/TL did not show a similar trend ($p = 0.79$). While the overall trend of RML volume loss and compensatory ΔRLL/RL volume increase was significant, pairwise comparisons reached statistical significance only between grades 0–1 and occasionally in grades 1–2, but not between higher grades (2–3 or 3–4), likely due to low sample sizes in those categories (Fig. 1A–E).

While individual comparisons between consecutive atelectasis grades did not reach statistical significance, comparisons after pooling the grades of atelectasis as minimal-subsegmental (grades 1–2) and segmental-lobar collapse (grades 3–4) showed a statistically significant difference (Table 4). Notably, changes in ΔRML, ΔRML/RL, and ΔRML/TL between grades 0 vs 1–2 and 1–2 vs 3–4 were highly significant ($p < 0.001$). And while ΔRLL/RL across atelectasis grades also reached significance, ΔRLL/TL did not (Fig. 1F–J).

Sensitivity analysis of segmentation accuracy and volumetric assessments after excluding RML wedge resections and replacing the original unedited outputs

Excluding cases with additional RML wedge resections beyond the index RULobectomy procedure, which could confound the interpretation of lobar volume loss, resulted in similar directional trends in volume metrics, reinforcing the association between automated volume metrics and physician-assigned atelectasis severity grades (Supplementary Fig. 2).

Moreover, analysis using the unedited segmentation outputs for the postoperative dataset yielded similar statistical differences in most volume metrics across both atelectasis and pooled atelectasis grades (Supplementary Fig. 3).

Table 3 | Quantitative assessment of lobar volume changes between preoperative and 6-month postoperative CT scans, stratified by radiographic atelectasis grade of 0–4

Metrics	None (0)	Minimal (1)	Sub-segmental (2)	Segmental (3)	Lobar (4)	*p*-value (for trend)
Median RML volume change (mL)	−4.6 (−78.5, 59.0)	−74.3 (−146.5, −40.5)	−189.9 (−296.2, −96.9)	−244.9 (−289.6, −199.2)	−317.8 (−440.7, −194.8)	<0.001
Median RML/RL volume change (%)	3.7 (1.1, 6.5)	0.7 (−2.3, 3.2)	−4.5 (−8.6, −1.1)	−9.2 (−12.1, −5.1)	−9.5 (−12.1, −7.0)	<0.001
Median RML/TL volume change (%)	0.6 (−0.9, 1.8)	−1.7 (−2.7, −0.2)	−3.1 (−5.2, −1.5)	−5.2 (−6.7, −3.5)	−5.1 (−6.4, −3.7)	<0.001
Median RLL/RL volume change (%)	31.2 (27.5, 36.3)	34.0 (29.4, 37.1)	38.2 (36.0, 42.5)	42.7 (35.6, 47.1)	50.5 (49.2, 51.9)	<0.001
Median RLL/TL volume change (%)	11.6 (9.5, 14.0)	11.1 (5.8, 13.4)	12.4 (9.1, 15.3)	13.0 (6.0, 19.5)	17.5 (17.1, 18.0)	0.79

RML right middle lobe, *RLL* right lower lobe, *RL* right lung, *TL* total lung.
Volume changes are presented as absolute (mL) and percentage changes (%) relative to right lung (RL) and total lung (TL) volumes (Median, IQR).

Qualitative analysis

Various grades of RML atelectasis are illustrated in Figs. 2 and 3. Qualitative error analysis on segmentation errors was performed on CT scans where the Dice score was <0.97. The most frequent errors were observed along the fissure between the RUL-RML and at the hilar junction of RML–RLL. Additional inaccuracies included over-segmentation of solid regions within dense consolidation areas and under-segmentation of bullae (Supplementary Fig. 4).

A detailed qualitative error analysis was performed on the external validation dataset, which is the most diverse publicly available dataset of chest CT scans for lung lobe segmentation. Representative errors over complex anatomical variations and challenging pathologies, such as scoliosis, chest wall deformities, severe COPD, large pleural effusions with near-complete lung collapse, and other abnormalities are illustrated in Supplementary Fig. 5.

Association with clinical outcomes

Correlation analysis between volume metrics and ΔPFT% revealed weak associations ($r < 0.3$) with RML volume metrics. However, ΔRLL/RL ($r = -0.43$, $p = 0.047$) and ΔRLL/TL ($r = -0.65$, $p = 0.001$) demonstrated a negative correlation with $\Delta FEV_1/FVC\%$ (Fig. 4).

Among 236 patients, 41 (17.3%) experienced the composite complication outcome within 1 year, comprising 23 (9.7%) readmissions, 19 (8.0%) bronchoscopies, and 6 (2.5%) reoperations. Among the bronchoscopies, 11/19 (57.8%) were clearly performed for the indication of RML collapse/atelectasis. The remainder of the RML atelectases were managed conservatively without bronchoscopic intervention.

Multivariable logistic regression showed no significant association between lobar volume metrics and 1-year composite clinical complications (Table 5). However, for bronchoscopy, ΔRML/RL [OR(95% CI): 0.89(0.81–0.98), $p = 0.01$] and ΔRML/TL [0.80(0.65–0.98), $p = 0.03$] were statistically significant predictors, indicating that lower proportional RML volume—a proxy for RML atelectasis—was associated with an increased likelihood of requiring bronchoscopic evaluation/intervention. (Table 6). Given this, we selected ΔRML/RL as the primary volume metric for subsequent evaluations involving airway metrics.

Airway segmentation, preoperative airway metrics, and their association with atelectasis

On the preoperative training dataset ($n = 34$), the airway model achieved a mean Dice of 0.94 ± 0.02. Internal ($n = 7$) and external ($n = 27$) test datasets yielded mean Dice scores of 0.95 ± 0.01 and 0.87 ± 0.03, respectively. On the postoperative training ($n = 59$) and internal test ($n = 22$) datasets, mean Dice was 0.95 ± 0.01 and 0.94 ± 0.02, respectively (Table 2).

The median(IQR) measures were as follows: RML bronchus length [manual-18 mm(16–20.5), automated-16.3 mm(14.1–19.1)], mean RML bronchus cross-sectional area [36.1 mm^2(28.8–42.7)], RML bronchus-intermedius angle [54°(49.1–58.9)], and RML bronchus angles with the planes [sagittal - -31.6°(-37.1, -26.7), coronal - 42.2°(36.5-47.1), transverse - -30.7°(-37, -22.9)]. The manually measured airway metric—RML bronchus length—demonstrated strong correlation with the automated RML bronchus length ($r = 0.71$, $p < 0.001$).

On multivariate linear and logistic regression, RML bronchus length was the only metric that was associated with ΔRML/RL volume [β(95% CI): 0.47(0.13–0.81), $p < 0.01$] (Table 7) and with physician-assigned atelectasis [OR(95% CI): 0.89(0.80–0.997), $p = 0.04$] (Table 8).

Discussion

In this study, we developed and validated an automated, deep learning-based volumetric pipeline to quantify and grade atelectasis–one of the most common clinical problems in pulmonary medicine and surgery. We selected postoperative atelectasis of the middle lobe following right upper lobectomy as a representative model to demonstrate that atelectasis can be graded using lobar volumetry. Importantly, such an automated system to reliably quantify atelectasis would also likely be

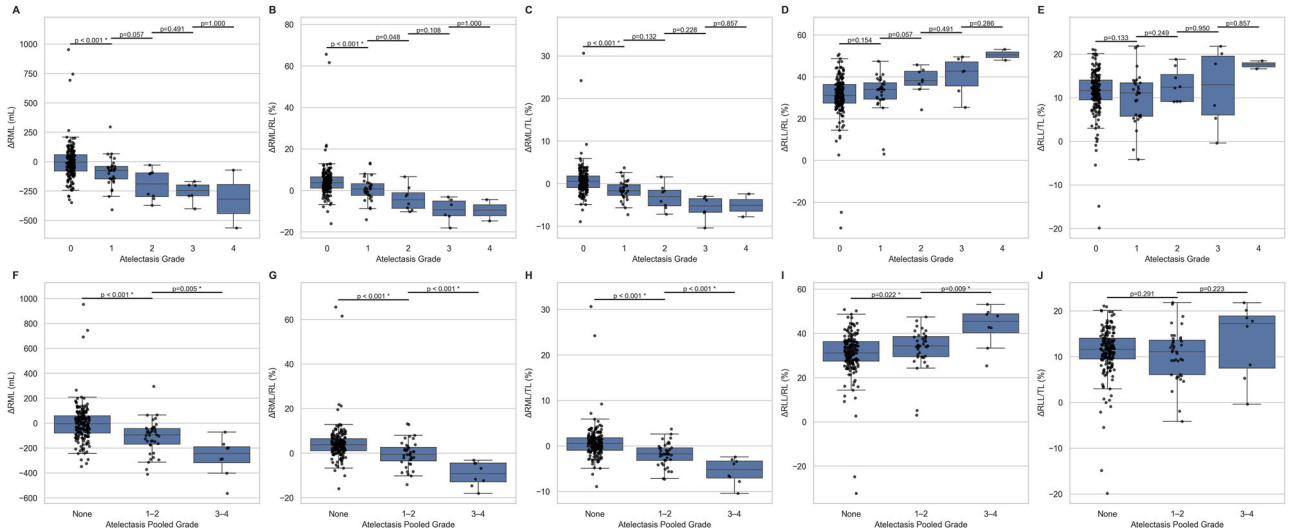

Fig. 1 | Boxplots showing changes in RML and RLL volume (ΔRML, ΔRLL) and their normalized ratios with respect to RL and TL volumes across increasing grades of atelectasis. The top row (**A**–**E**) shows the study cohort stratified into five atelectasis grades (0–4: none, minimal, subsegmental, segmental, and lobar). The bottom row (**F**–**J**) shows pooled atelectasis grades (none, minimal-subsegmental, segmental-lobar). Asterisks indicate statistically significant pairwise comparisons (with Bonferroni correction) (RML – right middle lobe, RLL -right lower lobe, RL right lung, TL – total lung).

Table 4 | Quantitative assessment of lobar volume changes between preoperative and 6-month postoperative CT scans, stratified by pooled atelectasis grades (grade 0, 1–2, and 3–4) Volume changes are presented as absolute (mL) and percentage changes (%) relative to right lung (RL) and total lung (TL) volumes (*p*-values for pairwise comparisons in Fig. 1F–J) (Median, IQR)

Metrics	None (0)	Minimal (1) and Sub-segmental (2)	Segmental (3) and Lobar (4)	*p*-value (for trend)
Median RML volume change (mL)	−4.4 (−78.6, 59.0)	−93.7 (−168.6, −43.7)	−244.9 (−318.0, −190.9)	**<0.001**
Median RML/RL volume change (%)	3.7 (1.1, 6.5)	−0.5 (−3.5, 2.6)	-9.2 (-12.9, -4.5)	**<0.001**
Median RML/TL volume change (%)	0.6 (-0.9, 1.8)	−1.7 (−3.1, −0.4)	-5.2 (-7.0, -3.3)	**<0.001**
Median RLL/RL volume change (%)	31.2 (27.5, 36.3)	34.3 (29.6, 38.5)	45.4 (40.2, 48.8)	**<0.001**
Median RLL/TL volume change (%)	11.6 (9.5, 14.0)	11.1 (6.1, 13.6)	17.2 (7.5, 18.9)	0.72

RML right middle lobe, *RLL* right lower lobe, *RL* right lung, *TL* total lung.

applicable to the broader assessment of atelectasis across diverse clinical scenarios.

Although deep learning–based models have demonstrated high accuracy in delineating lobar anatomy with Dice scores of 92–96%[19–22] (Table 9), they have been trained on preoperative anatomically normal lungs, with limited validation in postoperative settings where anatomical distortion is common. To date, only one study assessed lobar segmentation before and after various lobectomies, but without examining associations with clinical outcomes[23]. Our models achieved high accuracy in delineating lobar anatomy, even in the anatomically distorted postoperative setting, and further demonstrated that volumetric reduction of the RML with compensatory hyperinflation of RLL correlated with increasing clinically determined grades of atelectasis. These findings remained consistent in sensitivity analyses. Together, the results support feasibility of automated lobar volumetry as an objective surrogate for grading atelectasis severity, with the caveat that some cases were excluded and some required manual adjustment.

Beyond lobar volumetry, we identified that a longer preoperative RML bronchus length was independently associated with lower odds of physician-assigned atelectasis (OR < 1) and higher ΔRML/RL (β coefficient > 0), i.e. lesser degree of atelectasis. These findings suggest that a shorter RML bronchus, perhaps may bend more sharply for the same upward displacement of the RML into the empty postoperative RUL space, thereby in some way predisposing to bronchial angulation, kinking, and airflow obstruction. Certainly, however, this is purely theoretical, and further studies would be required to validate this.

Despite claims of superior performance of novel architectures—transformer-based and Mamba-based over CNNs for 3D medical segmentation, the CNN-based nnU-Net demonstrated the best Dice scores in a recent comprehensive benchmarking of six 3D medical datasets[24]. In addition to its robust and generalizable performance with relatively small datasets and its self-configuring architecture, the added advantages of lower VRAM consumption and shorter training times vs. other architectures influenced our choice of nnU-Net for model training. Table 9 lists other preoperative lung lobe segmentation models reporting Dice scores up to 0.99, but which lack validation on external datasets like LOLA11[18].

Moreover, our study demonstrates that deep learning models can robustly segment even complex, anatomically distorted postoperative lung anatomy following lobectomy. We have also demonstrated the ability of the models to generalize across vendors/scanners, acquisition protocols, and reconstruction kernels. While the postoperative model was generally robust, ~6% of scans required manual editing to ensure accurate volumetry—typically in cases with more severe anatomical distortion—reflecting a limitation to full automation. But this is consistent with a human-centered AI workflow in which the model provides high-quality segmentation for >90% of cases and substantially reduces manual burden, with limited human refinement needed only for a small subset and further performance gains expected with larger, more diverse training datasets.

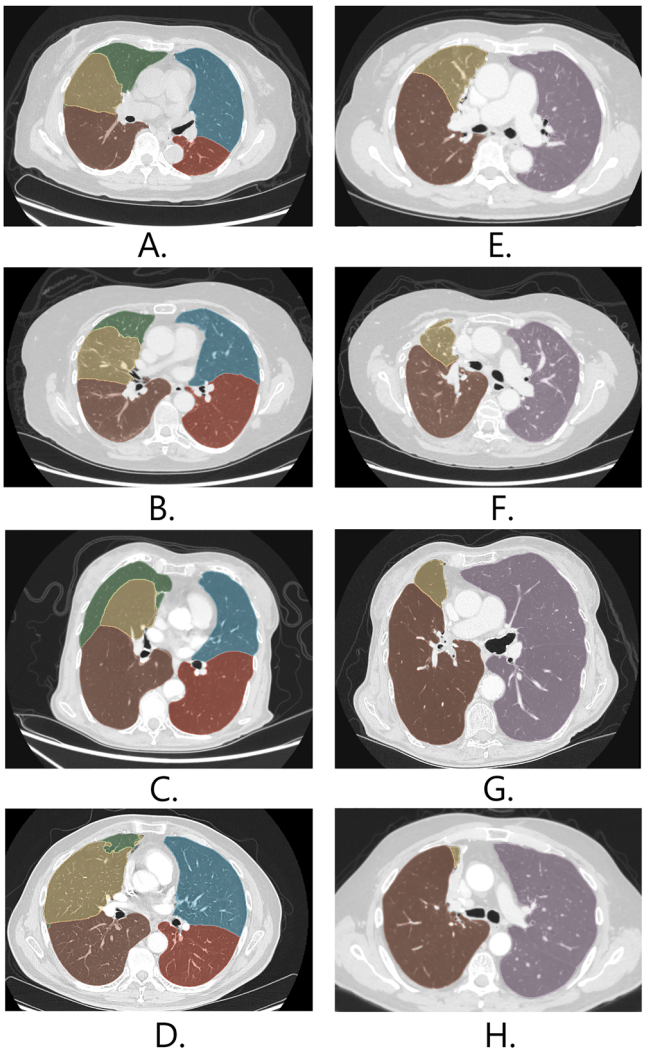

Fig. 2 | Axial CT slices showing preoperative (left column) and 6-month postoperative (right column) lobe segmentations in patients with increasing grades of atelectasis. The lobe in yellow in right column depicts the progressively increasing grades of atelectasis: minimal (grade 1) (**A**, **E**), subsegmental (grade 2) (**B**, **F**), segmental (grade 3) (**C**, **G**), and lobar atelectasis (grade 4) (**D**, **H**). Lobe labels (colors): RUL (green), RML (yellow), RLL (brown), LUL (blue), LLL (red), LL (pink) (RUL right upper lobe, RLL right lower lobe, RML right middle lobe, LUL left upper lobe, LLL left lower lobe).

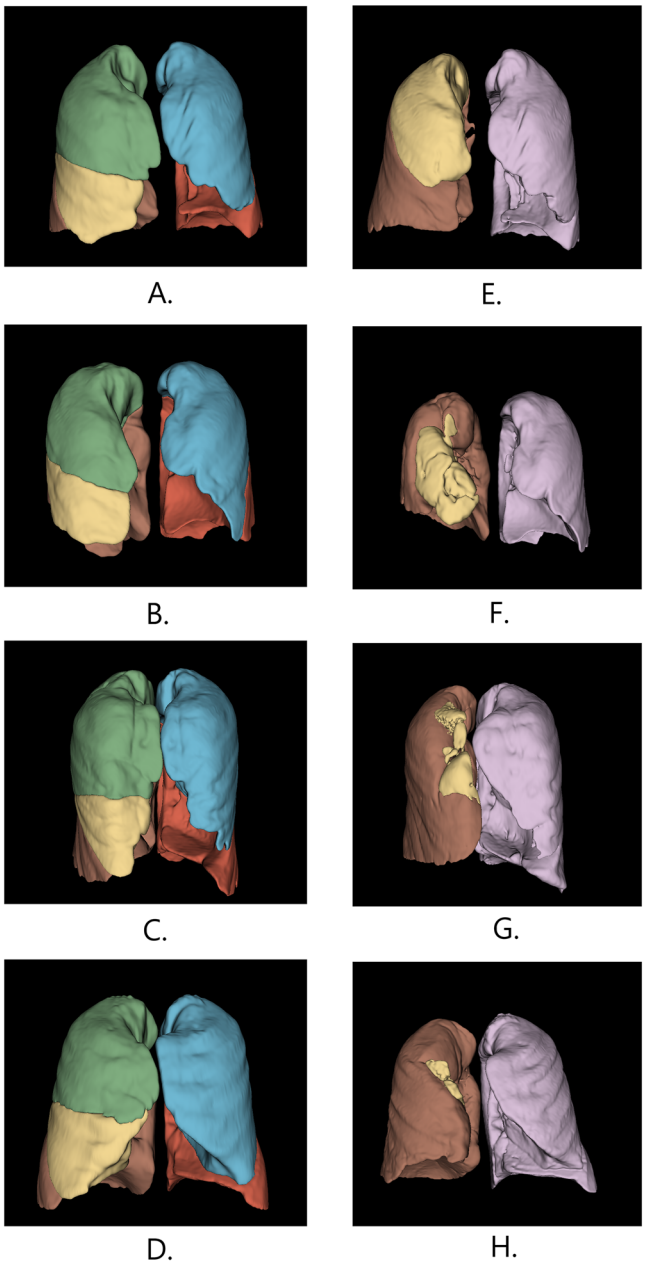

Fig. 3 | 3D modelling of lobe segmentations of preoperative (left) and 6-month postoperative (right) CT scans in patients with increasing severity of postoperative atelectasis. The lobe in yellow in the right column depicts the progressively increasing grades of atelectasis: minimal (**A**, **E**), subsegmental (**B**, **F**), segmental (**C**, **G**), and lobar atelectasis (**D**, **H**). Lobe labels (colors): RUL (green), RML (yellow), RLL (brown), LUL (blue), LLL (red), LL (pink) (RUL right upper lobe, RLL right lower lobe, RML right middle lobe, LUL left upper lobe, LLL left lower lobe).

To our knowledge, this is the first study to apply DL–based analysis for grading postoperative atelectasis. While prior work has attempted binary classification of atelectasis presence[25–27], such approaches are overly simplistic and fail to reflect the true spectrum of disease severity. Moreover, manual grading systems are prone to mild-moderate inter- and intra-rater variability, limiting their reliability and scalability. Our approach addresses these limitations by not only detecting presence of atelectasis as a "present/absent" decision, but also assessing its severity in a more objective, continuous, and reproducible manner by quantifying lobar volume loss. While our study focused only on RML atelectasis, this approach is readily applicable to other lobes and clinical scenarios where serial imaging is available.

It is important to consider a few clinical implications of this volumetric approach. While subtle streak-like atelectasis (grade 1) may not lead to overt volume loss, we observed significant differences in volume metrics between grades 0 and 1, suggesting that early changes can also be quantitatively captured. On a different note, our model segments lung lobes with all included structures, such as tumor masses or areas of consolidation. And notably, unlike atelectasis, tumor masses or consolidation from pneumonia typically do not lead to a substantial reduction in lobar volume. Therefore, the volumetric approach may offer the potential to differentiate atelectasis from other causes of increased opacity, such as pneumonia. This distinction, however, would require more detailed validation.

This volumetric approach for grading atelectasis also aligns with the principles of clinically explainable AI (xAI) as it yields outputs that are directly interpretable by clinicians, radiologists, or even technicians. This is unlike classification-based models, which often function as opaque "black boxes". Furthermore, developing a reliable classification model for atelectasis grading would require a much larger dataset to capture the wide

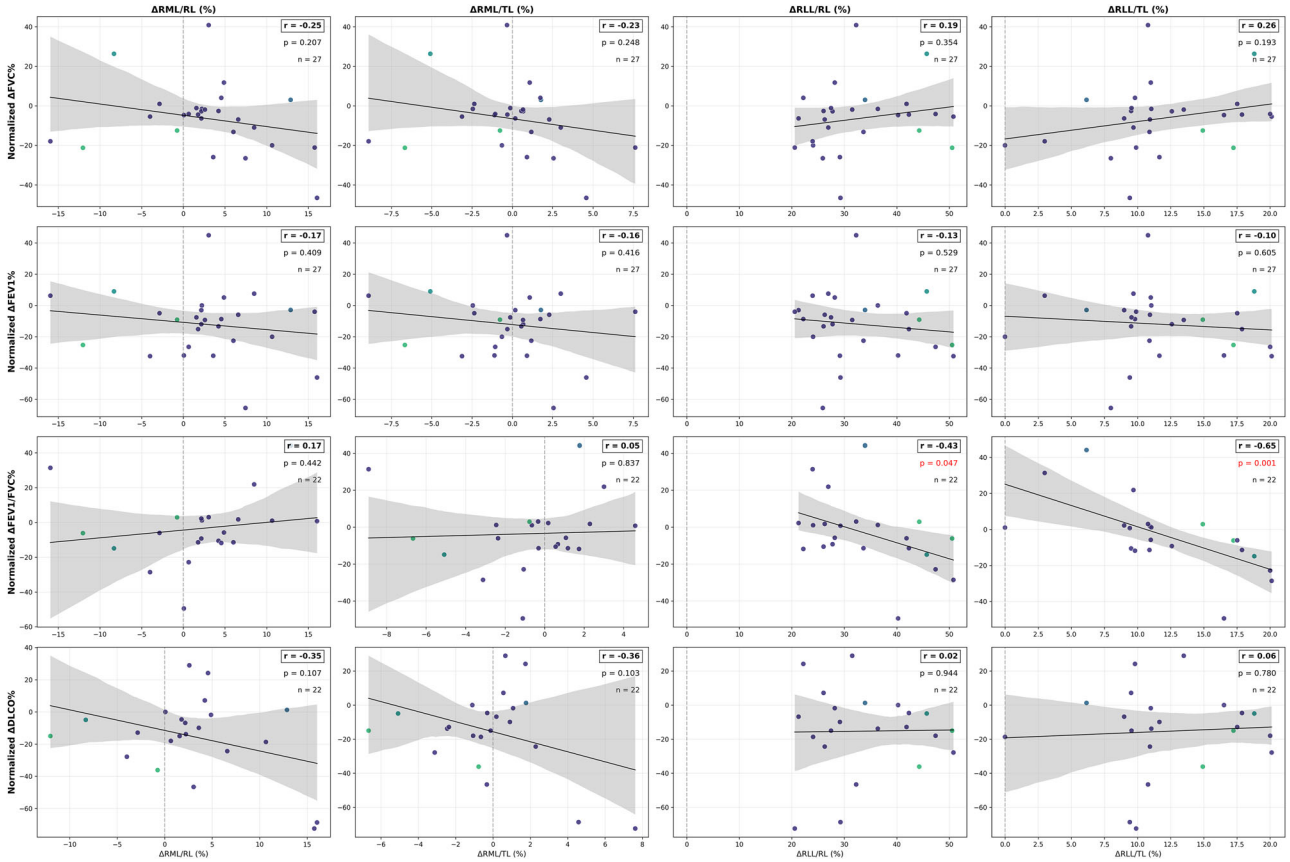

Fig. 4 | Correlation between lobar volume metrics and normalized changes in pulmonary function tests. Scatter plots showing associations between lobar volume metrics (ΔRML/RL, ΔRML/TL, ΔRLL/RL, ΔRLL/TL) at 6 months and normalized changes in pulmonary function tests (ΔFVC%, ΔFEV$_1$%, ΔFEV$_1$/FVC%, ΔDLCO%) at 3–24 months postoperatively. Points are colored by atelectasis severity grade (0–4). Pearson correlation coefficients (*r*), *p*-values, and sample sizes (*n*) are displayed for each comparison (RLL -right lower lobe, RML – right middle lobe, RL – right lung, TL – total lung, FVC – forced vital capacity, FEV$_1$ - forced expiratory volume in 1 second, DLCO - diffusing capacity of the lung for carbon monoxide).

Table 5 | Univariate and multivariable logistic regression analysis for 1-year composite outcomes

Variables		Univariate		Multivariate*	
		OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Age at surgery (years)		1.03 (1.00–1.07)	**0.04**		
Male Gender (vs. Female)		2.48 (1.26–4.95)	**0.01**	-	
Surgical approach (vs. Open)	VATS	0.50 (0.23–1.10)	**0.08**	1.26 (0.40–3.99)	0.70[a]
	Robot	0.52 (0.19–1.38)	**0.19**	1.69 (0.47–6.14)	0.42[a]
Congestive heart failure		1.81 (0.55–6.00)	0.33	-	
Chronic kidney disease		1.50 (0.56–4.01)	0.42	-	
Chronic obstructive pulmonary disease		1.85 (0.82–4.17)	**0.14**	-	
Myocardial infarction		2.82 (1.20–6.60)	**0.02**	-	
Peripheral vascular disease		1.96 (0.91–4.23)	**0.08**	-	
Length of stay (days)		1.15 (1.06–1.24)	**0.001**	1.17 (1.05–1.29)[a]	**0.003**[a]
Preoperative FEV$_1$ (%)		0.98 (0.96–0.99)	**0.007**	0.99 (0.97–1.01)[a]	0.59[a]
Volume metrics	ΔRML/RL (%)	0.98 (0.94–1.04)	0.58	0.96 (0.89–1.03)[b]	0.22[b]
	ΔRML/TL (%)	0.95 (0.85–1.06)	0.38	0.90 (0.78–1.01)[b]	0.18[b]
	ΔRLL/RL (%)	1.00 (0.97–1.04)	0.87	1.01 (0.96–1.05)[b]	0.76[b]
	ΔRLL/TL (%)	0.95 (0.90–1.01)	**0.12**	0.97 (0.90–1.04)[b]	0.37[b]

VATS video-assisted thoracoscopic surgery, *OR* odds ratio. *Only 4 variables were included given the 10 events per variable (EPV) rule.
[a]ORs and *p*-values are reported for the variables only for the model with ΔRML/RL.
[b]ORs and *p*-values are reported for separate multivariate models, each adjusting for length of stay, preoperative FEV$_1$, and surgical approach separately.

Table 6 | Univariate and multivariable logistic regression analysis for 1-year bronchoscopy

Variables		Univariate		Multivariate	
		OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Age at surgery (years)		1.04 (1.00–1.09)	**0.06**		
Male Gender (vs. Female)		1.94 (0.76–4.98)	**0.17**	-	
Surgical approach (vs. Open)	VATS	0.33 (0.12–0.94)	**0.04**	-	
	Robot	0.32 (0.08–1.28)	**0.11**	-	
Congestive heart failure		0.80 (0.10–6.48)	0.84	-	
Chronic kidney disease		1.58 (0.42–5.84)	0.49	-	
Chronic obstructive pulmonary disease		1.92 (0.65–5.69)	0.23	-	
Myocardial infarction		1.32 (0.36–4.83)	0.67	-	
Peripheral vascular disease		1.11 (0.35–3.52)	0.86	-	
Length of stay (days)		1.09 (1.02–1.17)	**0.01**	1.07 (0.97–1.18)[a]	0.19[a]
Preoperative FEV_1 (%)		0.97 (0.95–0.99)	**0.006**	0.97 (0.95–1.00)[a]	0.07[a]
Volume metrics	ΔRML/RL (%)	0.93 (0.86–1.1)	**0.08**	0.89 (0.81–0.98)[b]	**0.01**[b]
	ΔRML/TL (%)	0.86 (0.72–1.01)	**0.08**	0.80 (0.65–0.98)[b]	**0.03**[b]
	ΔRLL/RL (%)	1.03 (0.97–1.09)	0.39	1.02 (0.95–1.09)[b]	0.60[b]
	ΔRLL/TL (%)	0.98 (0.90–1.06)	0.57	0.98 (0.89–1.08)[b]	0.74[b]

VATS video-assisted thoracoscopic surgery, *OR* odds ratio. *Only 3 variables were included given the 10 events per variable (EPV) rule.
[a]ORs and *p*-values are reported for the two variables only for the model with ΔRML/RL.
[b]ORs and *p*-values are reported for their combination with length of stay and preoperative FEV_1 separately.

Table 7 | Univariate and multivariate linear regression to evaluate the association between airways metrics and the continuous outcome of ΔRML/RL volume change

Variables			Univariate		Multivariate	
			β coefficient (95% CI)	*p*-value	β coefficient (95% CI)	*p*-value
Age at surgery (years)			0.01(−0.09 to 0.11)	0.83		
Surgical approach (vs. Open)		VATS	−1.15(−3.34 to 1.04)	0.31		
		Robotic	3.13(0.60–5.67)	**0.02**	3.03(0.52–5.55)	**0.02**
Preoperative FEV_1 (%)			−0.01(−0.06 to 0.05)	0.83		
Charlson-Deyo score			0.23(−0.19 to 0.65)	0.28		
Diabetes mellitus			−0.50(−3.28 to 2.28)	0.72		
Myocardial infarction			1.40(−1.89 to 4.70)	0.40		
COPD			0.92(−2.04 to 3.88)	0.54		
Chronic kidney disease			0.74(−2.80 to 4.28)	0.68		
Airway metrics	RML bronchus length (mm) (manual)		0.43(0.08 to 0.78)	**0.01**	0.47(0.12–0.81)	**0.01**
	RML bronchus length (mm) (automated)		0.12(−0.12 to 0.36)	0.34		
	RML-Intermedius angle (°)		−0.16(−0.29 to −0.02)	**0.02**	−0.15(−0.28 to −0.01)	**0.03**
	RML transverse angle (°)		−0.06(−0.16 to 0.04)	0.26		
	RML sagittal angle (°)		0.04(−0.08 to 0.18)	0.47		
	RML coronal angle (°)		−0.05(−0.16 to 0.06)	0.35		
	Mean RML bronchus cross-sectional area (mm^2)		0.06(−0.04 to 0.16)	0.25		

variability in radiographic appearances, particularly subtle findings like linear opacities in grade 1—despite the positional invariance of CNNs. Hence, this segmentation-based method offers a more interpretable means of grading atelectasis and, by focusing solely on aerated volume, provides a broadly applicable solution across clinical contexts that remains effective even with limited training data.

This study offers several key strengths. Firstly, it leverages a large real-world clinical cohort for automated quantification of atelectasis, using clinically interpretable AI-derived volumetry. By validating model outputs against physician-assigned atelectasis grades and demonstrating ΔRML/RL and ΔRML/TL as potential markers of need for clinical intervention, such as bronchoscopy, the study exemplifies clinically applied AI, in contrast to much of AI research in healthcare that is decoupled from real-world outcomes. Unlike black-box classification models, our pipeline generates outputs that can be reviewed and validated by clinicians/technicians in real time, exemplifying verifiable AI.

Table 8 | Univariate and multivariate logistic regression to evaluate the association between airways metrics and the categorical outcome of RML atelectasis on CT at 6–9 months

Variables			Univariate		Multivariate	
			OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Age at surgery (years)			0.99(0.96–1.02)	0.52		
Surgical approach (vs. Open)		VATS	1.41(0.73–2.72)	0.31		
		Robotic	0.75(0.34–1.68)	0.49		
Preoperative FEV_1 (%)			0.99(0.98–1.01)	0.52		
Charlson-Deyo score			1.03(0.91–1.16)	0.66		
Diabetes mellitus			1.26(0.57–2.79)	0.56		
Myocardial infarction			1.29(0.51–3.23)	0.51		
COPD			0.99(0.42–2.33)	0.99		
Chronic kidney disease			1.20(0.45–3.20)	0.71		
Airway metrics	RML bronchus length (mm) (manual)		0.88(0.80–0.98)	**0.02**	0.89(0.80–0.997)	**0.04**
	RML bronchus length (mm) (automated)		1.01(0.95–1.08)	0.73		
	RML-Intermedius angle (°)		1.01(0.97–1.05)	0.58		
	RML transverse angle (°)		1.01(0.98–1.04)	0.47		
	RML sagittal angle (°)		0.99(0.96–1.03)	0.74		
	RML coronal angle (°)		0.99(0.96–1.02)	0.54		
	Mean RML bronchus cross-sectional area (mm^2)		0.96(0.93–0.996)	**0.03**	0.97(0.94–1.00)	**0.05**

Table 9 | Prior state-of-the-art lung lobe segmentation performance reported in the literature and on the publicly available LOLA11 challenge dataset comprising 55 chest CT scans

Author (Year) [citation]	Right upper lobe	Right middle lobe	Right lower lobe	Left upper lobe	Left lower lobe	Overall
Doel et al. (2012) – Traditional CV methods[34]	0.86 ± 0.15	0.55 ± 0.40	0.77 ± 0.34	0.88 ± 0.21	0.86 ± 0.24	0.79 ± 0.31
3D U-Net (2016-2020)[35–37]	0.93 ± 0.07	0.84 ± 0.12	0.93 ± 0.04	0.92 ± 0.04	0.93 ± 0.04	0.91 ± 0.08
Ferreira et al. FRV-Net (2018)[38]	0.94 ± 0.07	0.87 ± 0.12	0.94 ± 0.05	0.95 ± 0.03	0.95 ± 0.04	0.92 ± 0.02
Gerard et al. Series of 3D CNN (2019)[39]	0.99	0.98	0.99	0.99	0.99	0.99
Xie et al. RU-Net (2020)[21] – (IoU)	0.95 ± 0.03	0.96 ± 0.03	0.96 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.95 ± 0.03
Zheng et al. (2021) - Dual attention network[40]	0.93 ± 0.01	0.90 ± 0.01	0.96 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.93 ± 0.02
Zhang et al. (2021)[41] DenseVNet	0.96	0.92	0.96	0.94	0.93	0.94
Peng et al. (2022)[22] - Multi-feature fusion and Ensemble.	0.97	0.88	0.96	0.98	0.97	0.95
Bao et al. (2023)[42] Edge enhancement cascaded network	0.98	0.96	0.98	0.98	0.97	0.97
Nomura et al. (2025)[43] TriSwinUNETR	0.93	0.85	0.95	0.97	0.96	0.93
LOLA11 lung lobe segmentation challenge[18]						
Van Rikxkoort et al. (2010)[44] Traditional CV						0.85
Lassen et al. (2013)[45] Traditional CV						0.88
Bragman et al. (2017)[19] Probabilistic model	0.91 ± 0.20	0.88 ± 0.24	0.93 ± 0.07	0.80 ± 0.23	0.91 ± 0.19	0.88
Imran et al. (2018)[20] PDV-Net	0.93 ± 0.07	0.86 ± 0.12	0.95 ± 0.03	0.93 ± 0.03	0.94 ± 0.03	0.92 ± 0.07
Current SOTA – nnU-Net (2025)	0.95 ± 0.08	0.86 ± 0.22	0.96 ± 0.06	0.95 ± 0.13	0.92 ± 0.19	**0.928 ± 0.15**
Ours (2025)	**0.95 ± 0.06**	**0.86 ± 0.22**	0.94 ± 0.13	**0.95 ± 0.13**	0.91 ± 0.17	**0.923 ± 0.16**

Its utility is further enhanced by the availability of open-source scripts for verifying segmentation accuracy and allowing integration into the backend of most software. A major technical strength is the use of nnU-Net—a widely validated segmentation framework in biomedical imaging—requiring relatively lesser expertise to deploy. Additionally, the volume metrics enable atelectasis to be measured as a continuous variable, preserving more information than categorical grading, allowing for more nuanced statistical modeling. Moreover, continuous outcomes also provide greater statistical power and enable smaller sample sizes in clinical studies, supporting their use as efficient and meaningful endpoints in clinical trials.

This work offers several potential avenues for both research and clinical application. From a research perspective, we have already applied this framework in a retrospective study (under review) assessing whether surgical fixation ("pexy") of the RML reduces postoperative RML atelectasis and torsion[28]. Beyond this, the model outputs of lobar volumes/atelectasis could

be used as quantitative endpoints in clinical trials focused on ventilation strategies and other interventions to prevent atelectasis, or on respiratory biomechanics across anesthesiology, critical care, and thoracic surgery. Clinically, the lobe segmentation model could be integrated into workflows such as preoperative planning of lung surgery and tumor localization through lobar mapping, along with other models of tumor segmentation. It may also serve as an objective tool for postoperative assessment of atelectasis in research or clinical applications. These applications, however, warrant validation in prospective, context-specific studies.

Despite the strengths of this study, several limitations must be acknowledged. While the preoperative model demonstrated strong generalizability, including external validation, caution is warranted in patients with chest wall or vertebral anomalies, and at the RUL/RML junction, where fissure incompleteness is often observed. Additionally, errors (Supplementary Figs. 4 and 5) may stem from inconsistencies in reference standard annotations, a persistent challenge in biomedical AI. The postoperative model lacks external validation due to the absence of publicly available datasets. Furthermore, the clinical outcome of atelectasis grading was derived from a single-center retrospective cohort, which may limit generalizability. Our thoracic surgery center, for example, cares for a particularly healthy population of non-smokers of Asian ancestry who develop lung cancer as a result of a genetic predisposition. These patients may have different predispositions to and/or effects from RML atelectasis than the sicker, tobacco-addicted population with COPD typically seen at other centers. There may also be substantial, purely racial differences between our population and those treated at other centers. Hence, until larger, more diverse datasets become available, applying automated lobe volumes for atelectasis grading may require visual confirmation. Special consideration is needed in cases with pleural pathology or external compression, which can globally affect lobar volumes and misrepresent aeration and relative lobar volumes. Importantly, the current pipeline is designed for longitudinal comparison and cannot interpret a single CT scan in isolation. Additionally, although we demonstrated the utility of continuous volume metrics as a surrogate for atelectasis, we did not define specific thresholds for grading atelectasis—an area that warrants further investigation and validation in future studies with clinical measurements like arterial blood gases and PFTs, to establish meaningful thresholds for clinical classification. Lastly, among the airway metrics introduced, only one underwent manual validation, whereas others were evaluated as exploratory, hypothesis-generating indices, highlighting the need for more detailed and robust evaluation of these automated airway metrics in future studies. We developed and validated a deep learning–based volumetric pipeline to quantify postoperative atelectasis after right upper lobectomy. Using automated lobar segmentation and validation against physician-assigned grades of atelectasis, we show that right middle lobe volume loss serves as an objective and reproducible surrogate for differentiating clinical atelectasis severity. Among the evaluated volume metrics, ΔRML/RL and ΔRML/TL served as candidate markers for bronchoscopy requirement, indicating potential clinical relevance that warrants further exploration in larger cohorts. These findings support the broader applicability of deep learning–driven lobar segmentation and volumetry as a reliable method for grading postoperative atelectasis in both clinical and research settings, and potentially for any atelectasis across other types of applications in pulmonary medicine and thoracic surgery.

Methods

Study design and reporting standards

This study was a single-center retrospective analysis. The Checklist for Artificial Intelligence in Medical Imaging (CLAIM-2024) was used in study design and implementation[29]. The CLAIM checklist is modeled after the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines and is part of the EQUATOR network, ensuring best practices in reporting and facilitating the translation of AI into clinical practice. The study was approved by the Stanford University School of Medicine Institutional Review Board (IRB), which waived the requirement for informed consent as the study involved only retrospective chart review (IRB-70048). All procedures were carried out in accordance with the Declaration of Helsinki ethical standards.

Patient cohort and clinical data for inference

We retrospectively reviewed all consecutive patients who underwent a right upper lobectomy in the Division of Thoracic Surgery at Stanford University Hospital from January 2008 through December 2023. Patients were identified using Stanford's customized version of the Society of Thoracic Surgery (STS) General Thoracic Database and the STAnford Research Repository (STARR) clinical data warehouse. Patients without a preoperative CT scan within 6 months prior to surgery or a postoperative CT scan performed at approximately 6 months (range: 3–9 months) after surgery were excluded. Patients with scans of axial slice thickness >3 mm were also excluded. Baseline demographic and clinical data were extracted from the databases and the electronic medical record. Clinical follow-up data were collected for up to 12 months post-operatively. Out of the 438 patients who underwent right upper lobectomy during the study period, 236 patients were included in the study after applying the stated inclusion and exclusion criteria.

Imaging acquisition

Preoperative and postoperative diagnostic chest CT scans were acquired using multidetector CT scanners from multiple vendors, most commonly Siemens (Siemens Healthineers, Erlangen, Germany) and GE Medical Systems (GE Healthcare, Chicago, IL, USA), occasionally with Toshiba (Toshiba Medical Systems Corporation, Otawara, Japan) and Philips (Philips Healthcare, Best, The Netherlands). Scans were performed during end-inspiratory breath-hold when tolerated. A wide variety of reconstruction kernels was used across the cohort, with the most frequent being STANDARD, T20f, Tr20f, and T20s. Axial images were reconstructed with slice thicknesses ranging from 0.3 mm to 3 mm.

CT acquisition protocols varied, with tube voltages ranging from 80 to 130 kVp (median [IQR]: preoperative – 120 [120]; postoperative – 120 [80–120]) and tube currents ranging from 10 to 660 mA (median [IQR]: preoperative – 30 [20–35]; postoperative – 35 [20–35]). Most preoperative and postoperative scans were performed without intravenous contrast.

Training data

The training dataset was divided into training and test sets, ensuring that no patient had scans in both sets. A separate validation set was not required, as the self-configuring nnU-Net framework does not rely on manual hyperparameter tuning. The training pipelines for both the preoperative and postoperative segmentation models are illustrated in Fig. 5. Conversion of DICOM files to NIfTI format de-identified the images by removing all protected health information from the image metadata.

To ensure generalizability, the preoperative lobe segmentation model was trained using two datasets: an internal institutional dataset and an external public dataset (OttawaChestCT dataset). The internal Stanford dataset included a random subset of 41 CT scans (34 for training and 7 for test) selected from a total of 236 clinical scans. The OttawaChestCT dataset contributed 81 scans (59 for training and 22 for test) out of the original 100; 19 scans were excluded due to incomplete lobe annotations (i.e., missing one or more of the five lobe masks). In total, the combined dataset comprised 122 annotated CT scans, with voxel-wise labels for six classes: background, right upper lobe (RUL), RML, right lower lobe (RLL), left upper lobe (LUL), and left lower lobe (LLL).

The postoperative lobe segmentation model was trained using only the internal institutional dataset, consisting of randomly selected 82 CT scans (65 for training and 17 for test) from the full cohort of 236 clinical scans. Each scan comprised voxel-wise labels for four classes: background, RML, RLL, and left lung.

An airway segmentation model was trained using the same 41 preoperative and 82 postoperative CT scans from the internal Stanford dataset (101 for training and 22 for test). The AeroPath[30] dataset ($n = 27$) was used as the external test dataset. Each scan comprised voxel-wise labels for two classes: background and airway.

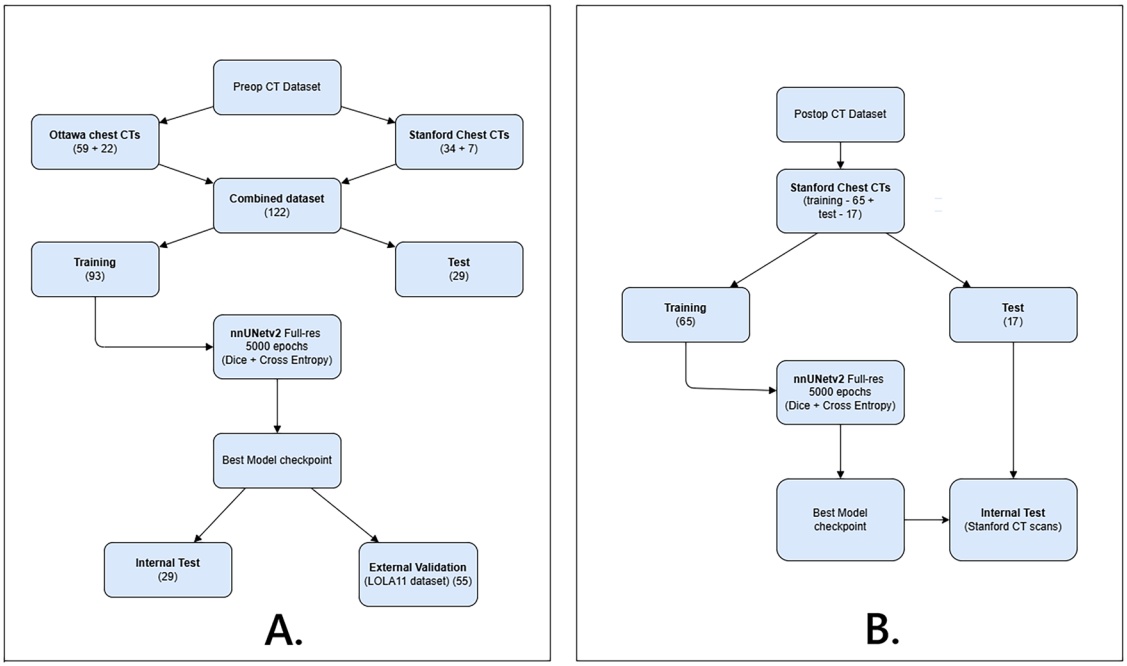

Fig. 5 | Overview of the model development pipelines using preoperative and postoperative CT datasets. A Preoperative CT model development: a total of 122 scans from Ottawa (59 training + 22 test) and Stanford (34 training + 7 test) chest CTs were combined. The dataset was split into 93 training and 29 validation scans, used to train a 3D nnUNetv2 full-resolution model for 5000 epochs with Dice +Cross Entropy loss. The best-performing model checkpoint was evaluated on both the internal test set ($n = 29$) and an external test set (LOLA11 dataset, $n = 55$). **B** Postoperative CT model development: a separate cohort of 82 Stanford postoperative chest CTs (65 for training, 17 for testing) was used to train a second 3D nnUNetv2 model using the same settings.

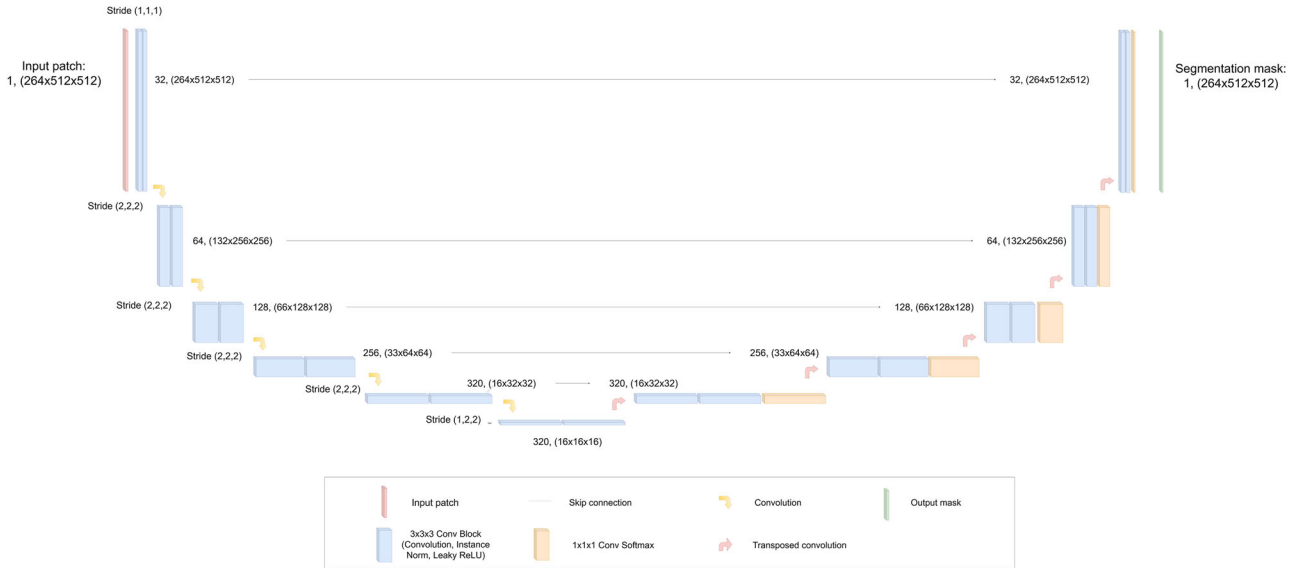

Fig. 6 | A 3D U-Net architecture for airway and lung-lobe segmentation consisting of six encoder stages with feature channels (32, 64, 128, 256, 320, 320) and five decoder stages, all using 3 × 3 × 3 3D convolutional kernels. The input shape 264 × 512 × 512 reflects the median CT volume size of the preoperative cohort and is shown for dataset context and representative scale only. Although the network architecture is fully convolutional, both training and inference use memory-efficient 3D patches processed via a sliding-window approach, with patch-level predictions aggregated to generate full-volume segmentations at inference.

Reference standard for segmentation masks

Reference standard segmentation masks for the lobe models were generated using 3D Slicer (version 8.5.1; MIT, Massachusetts, USA)[31] by D.N.K. for preoperative and postoperative CT scans in the internal Stanford dataset, while the original annotations provided by the dataset creators were used for the OttawaChestCT public dataset[32]. Reference standard segmentation masks for the airway model were generated using a validated airway segmentation model[30].

Model architecture

This study utilized the nnU-Net v2 model, a self-configuring deep learning framework for biomedical image segmentation. Its predecessor, nnU-Net v1, was built upon the U-Net architecture and designed to automatically adapt its network architecture, preprocessing steps, and training pipeline to the specific characteristics of a given dataset—eliminating the need for manual tuning of architectural or hyperparameter settings. The nnU-Net v2 retains the original self-adapting architecture and automated training

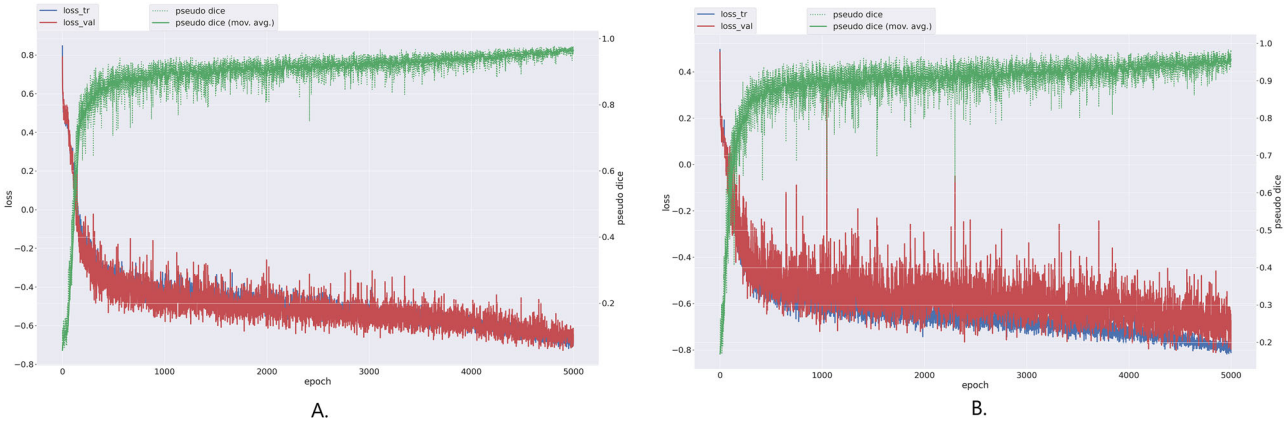

Fig. 7 | Training and validation performance of the segmentation models across epochs. Training and validation loss curves along with pseudo-Dice similarity coefficient (DSC) across epochs for the preoperative **A** and postoperative **B** models. In both panels, the blue line represents training loss, red line represents validation loss, dotted green line represents raw pseudo-Dice, and solid green line represents its moving average.

configuration of the nnU-Net framework, with enhancements focused on improved training efficiency and broader device compatibility. Detailed implementation of the model is presented in the original paper[24,33].

The lobe and airway segmentation network architectures followed the default 3D full-resolution Conv-UNet design, composed of a six-stage encoder (32, 64, 128, 256, 320, 320 features per stage) and a symmetric five-stage decoder, both using 3 × 3 × 3 Conv3D kernels (Fig. 6)

Training hyperparameters

A leaky ReLU activation function was used in all the hidden layers. The model was trained using the Adam optimizer with an initial learning rate of 0.001, followed by a polynomial learning rate decay schedule throughout training. Combined Dice and cross-entropy loss was used to guide optimization. Deep supervision was enabled. The checkpoint with the highest pseudo-Dice coefficient—i.e., the approximate Dice score computed on training patches rather than on the full CT volume—across all lung lobes on the training dataset was selected for further evaluation on both the test set and the full clinical dataset.

Training pipeline

Model training was performed using a single NVIDIA H100-SXM5 GPU. The preoperative and postoperative CT lobe segmentation models were trained for 5000 epochs each (Fig. 7), and the airway segmentation model was trained for 1000 epochs. Training time per epoch ranged from 22 to 30 s, resulting in a total training duration of approximately 42 h for each lobe model and 7 h for airway model.

Input CT images were resampled to a median voxel spacing of 1.25 × 0.70 × 0.70 mm for preoperative dataset, 1.0 × 0.70 × 0.70 mm for postoperative dataset, and 1.0 × 0.70 × 0.70 mm for airway dataset. These CT images were normalized using CT-specific intensity scaling. Due to the large volumetric size of CT scans in the dataset (median shape: 264 × 512 × 512—preoperative CTs, 314 × 512 × 512—postoperative CTs, 312 × 512 × 512—airway CTs), training was performed with a patch size of 96 × 160 × 160 voxels and a batch size of 2.

Preprocessing, data augmentation, and postprocessing

These were handled using the default nnU-Net pipeline. Preprocessing included resampling CT scans to the median dataset voxel spacing, intensity normalization using foreground intensity statistics, and cropping to nonzero regions to remove background. Data augmentation was performed on-the-fly and included random spatial transformation (rotation, scaling, elastic deformation) and intensity augmentations (gamma, noise, blur). Postprocessing involved removing all but the largest connected component for each predicted lobe label, applied selectively based on whether it improved the Dice score compared to baseline segmentation.

In addition to the default nnU-Net postprocessing, customized automated postprocessing was applied to further refine the predicted masks. This included removal of anatomically discordant contralateral lobe predictions and isolated 3D islands disconnected from the largest lobar component. These customized postprocessing steps were not part of the training pipeline but were implemented post-inference and resulted in a marginal improvement in overall Dice scores.

Due to the absence of reference standard masks for the full clinical dataset, predicted lobe segmentations were visually inspected to confirm adequate anatomical accuracy.

Outcomes—atelectasis severity

Atelectasis of the RML was graded using a 5-point scale: 0=none, 1=minimal/linear, 2=subsegmental, 3=segmental, and 4=near-total or total lobar collapse. CT scans were reviewed using a lung window setting (window width: 1500 HU; window level: –600 HU). Image evaluation was performed using 3D Slicer (version 8.5.1; MIT, Massachusetts, USA)[31]. Two board-certified surgeons independently reviewed and graded all scans, resolving discrepancies through consensus until full agreement was achieved.

Outcomes—clinical and functional outcomes of atelectasis

We evaluated 1-year clinical complications and functional outcomes to assess the predictive ability of our automated volume metrics for clinical events related to RML atelectasis following RULobectomy.

We evaluated complications over 1-year period, as RML atelectasis after RULobectomy (apart from the rare event of acute torsion) is a chronic process that typically develops and progresses for several months after surgery. The outcomes included were: (1) bronchoscopy, (2) readmissions, (3) reoperations, and (4) a composite outcome combining all three of these outcomes (i.e., having suffered any of these three outcomes).

For the functional outcomes, pulmonary function tests (PFTs) were assessed preoperatively (within 1 year before surgery) and postoperatively (3 months-2 years after surgery). We analyzed changes in %FVC, %FEV_1, %FEV_1/%FVC, and %DLCO, calculating absolute change (Δ = postoperative – preoperative) and normalized percentage change:

$$\text{Normalized}\Delta\text{PFT}\% = [(\text{postoperative}\%\text{PFT} - \text{preoperative}\%\text{PFT})/\text{preoperative}\%\text{PFT}] \times 100$$

Lobe model-based lobe volume metrics

Preoperative and postoperative lung volumes for the full clinical dataset were quantified using automatically segmented masks derived from lobe model inference on the respective CT scans. The preoperative CT segmented regions included RUL, RML, RLL, LUL, and LLL, while the

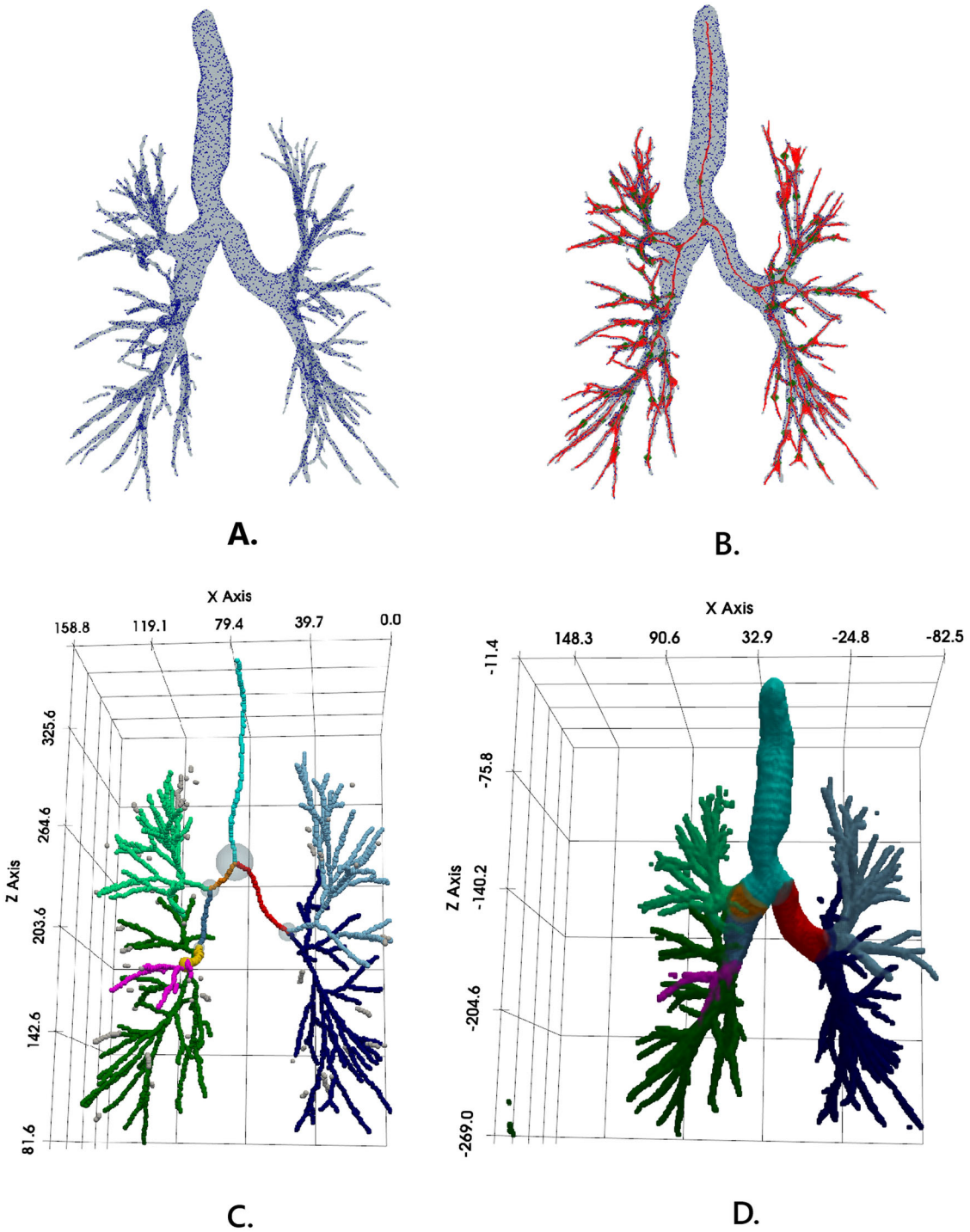

Fig. 8 | Sequential airway postprocessing workflow following automated CT airway segmentation. A Reconstructed airway up to 5th generation, visualized as blue surface vertices. **B** Airway centerline extraction via 3D skeletonization (red line), overlaid with detected bifurcations (red diamonds). **C** Anatomical labeling of airway using heuristics guided by detected airway bifurcations, enabling lobar airway classification (color key shown). **D** Airway segmented into different anatomic segments by nearest neighbor approach along the labeled skeleton. Color key: trachea (cyan); right main stem bronchus (orange); left main stem bronchus (red); bronchus intermedius (steel blue); RUL airway (light green); RLL airway (dark green); RML airways (magenta); RML bronchus (gold); LUL airways (sky blue); LLL airway tree (navy) (RUL – right upper lobe, RLL -right lower lobe, RML – right middle lobe, LUL left upper lobe, LLL – left lower lobe).

postoperative regions included RML, RLL, and left lung. Voxel volumes were calculated using the dimensions of each voxel. Lobar and total lung volumes were computed by multiplying the voxel volume (converted from mm^3 to mL) with the number of voxels in each label. A subset of the calculated volumes was cross validated against measurements obtained using the standard volume computation tool in 3D Slicer (version 8.5.1; MIT, Massachusetts, USA) to ensure accuracy. RML and RLL volumes were also normalized with right lung and total lung volumes to minimize inter-scan variability and reported as ΔRML/RL, ΔRML/TL, ΔRLL/RL, and ΔRLL/TL.

These ΔVolume Ratios were calculated as:

[Postoperative Volume Ratio – Preoperative Volume Ratio]

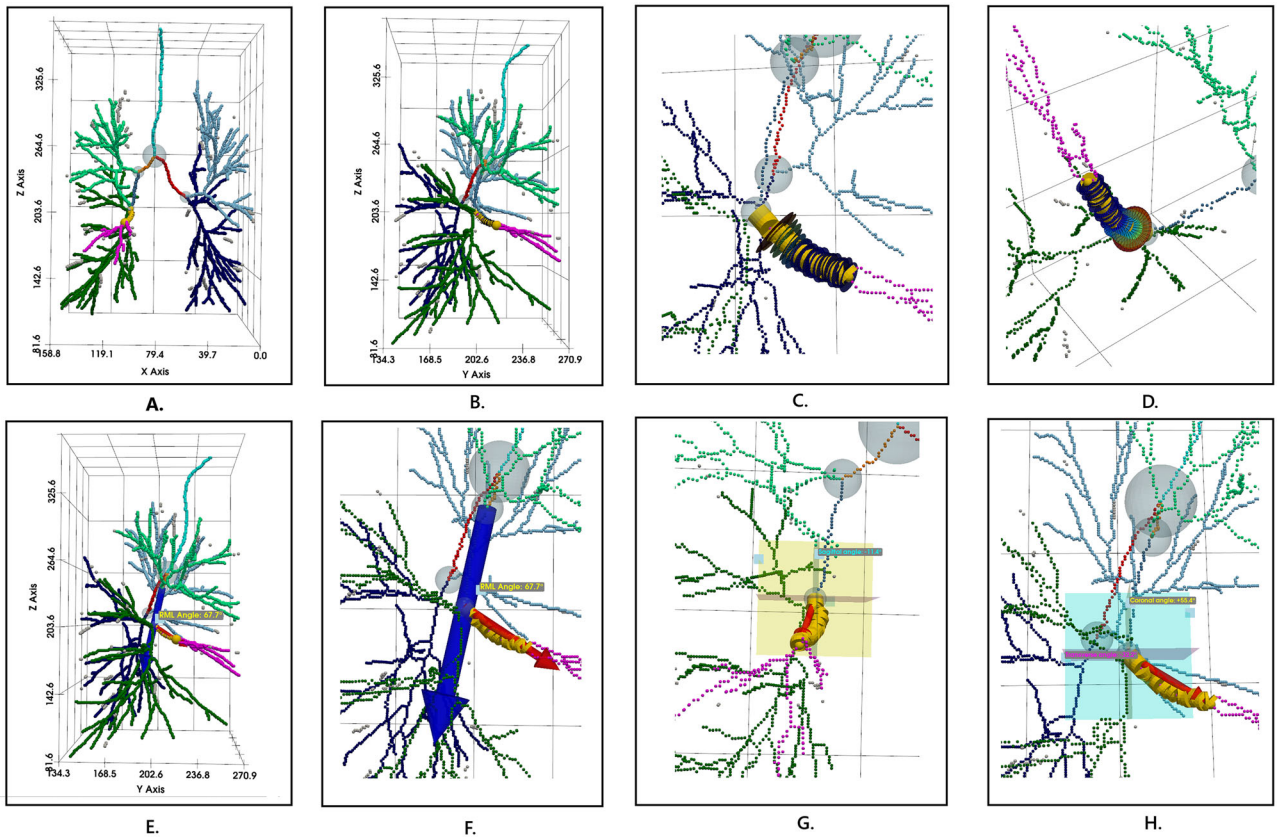

Fig. 9 | Automated airway metrics analysis from a single postoperative CT. Reconstructed airway tree shown from anterior (**A**), right-lateral (**B**), and left-superior (**C**) views, with the RML lobar bronchus highlighted (gold) and sampled for cross-sectional area (CSA) along its centerline (CSA depicted as discs) (**C**, **D**). **E**, **F** RML-bronchus intermedius angulation measured between the intermedius vector (blue arrow) and the RML bronchus vector (red arrow) (here 67.7°). **G**, **H** RML bronchus orientation relative to anatomical planes: sagittal (−11.4°), coronal (+55.4°), and transverse (−32.2°). Color key (applies across panels): trachea (cyan), right main stem bronchus (orange), bronchus intermedius (steel blue), RML bronchus (gold) (RML – right middle lobe).

Airway model-based airway metrics

Anatomical characteristics of preoperative airways for the full clinical dataset were detected using the trained airway segmentation model. These airway masks were skeletonized, and lobar airways were characterized using the bifurcations in the airways (Fig. 8). The RML bronchus was confirmed visually in all CT scans and several automated airway metrics were calculated such as length (RML bronchus length in mm), cross-sectional area (mean RML bronchus cross-section area in mm^2), and 3D orientation (angle between RML bronchus and bronchus intermedius, transverse, sagittal, and coronal planes) (Fig. 9)—any metrics that we felt may potentially have an impact upon the development of postoperative atelectasis.

Statistical analysis

Descriptive statistics were used to summarize baseline demographic and clinical data. The normality of continuous data was evaluated using the Shapiro–Wilk test. Normally distributed continuous variables were reported as mean ± standard deviation (SD) and compared using unpaired *t*-tests, while non-normally distributed variables were reported as median (interquartile range, IQR) and compared using the Mann–Whitney *U* test. Categorical variables were presented as frequencies and percentages and compared using Pearson's chi-square test (χ^2) or Fisher's exact test, as appropriate. The Jonckheere–Terpstra test was used to assess for ordered trends in all volume metrics across atelectasis severity grades.

The association between lobar volume metrics and clinical outcomes was assessed using univariate and multivariate logistic regression for binary outcomes (composite, bronchoscopy) and Pearson correlation for continuous outcomes (normalized Δ%PFTs). Given the limited number of events in the clinical outcomes and the 10 events per variable (EPV) requirement, we limited the modelling to only three covariates: length of hospital stay, preoperative FEV_1%, and one of the volume metrics. We selected length of hospital stay as a proxy for early postoperative status that may affect the 1-year outcomes, and we selected a single preoperative variable, FEV_1, as a proxy for preoperative respiratory function and forced the volume metrics into the model. The effect sizes were reported as odds ratios (ORs) with 95% confidence intervals (CIs).

The association between airway metrics and RML atelectasis was assessed by using univariate linear regression for the continuous outcome (ΔRML/RL) and univariate logistic regression for the binary outcome (physician-assigned atelectasis). Multivariable regression was subsequently performed using all covariates meeting eligibility ($p < 0.20$). The effect sizes (with 95% CIs) were reported as β coefficients for linear regression and ORs for logistic regression.

A two-tailed $\alpha < 0.05$ was considered statistically significant. Bonferroni correction was applied for multiple testing. All analyses were conducted using SPSS software (version 29.0, IBM Corp., Armonk, NY, USA), R (version 4.4.2, R Foundation for Statistical Computing, Vienna, Austria), and Python (version 3.13, Python Software Foundation, Wilmington, DE, USA).

Data availability

The public datasets used for preoperative lobe segmentation and airway segmentation are available in cited sources. However, the institutional dataset used for preoperative and postoperative lobe segmentation is not available due to institutional review board restrictions and patient privacy protections.

Code availability

The code and the final preoperative and postoperative model checkpoints for lung lobe and airway segmentation are available on Github [https://github.com/Devanish31/atelectasis] and Figshare [https://doi.org/10.6084/m9.figshare.29877509], respectively.

Received: 2 October 2025; Accepted: 17 April 2026;
Published online: 30 April 2026

References

1. Zeng, C., Lagier, D., Lee, J. W. & Vidal Melo, M. F. Perioperative pulmonary atelectasis: part I. Biology and mechanisms. *Anesthesiology* **136**, 181–205 (2022).
2. Tanner, T. G. & Colvin, M. O. Pulmonary complications of cardiac surgery. *Lung* **198**, 889–896 (2020).
3. Zhao, Y., Zheng, R., Xiang, W., Ning, D. & Li, Z. Systematic review and meta-analysis on perioperative intervention to prevent postoperative atelectasis complications after thoracic surgery. *Ann. Palliat. Med.* **10**, 10726–10734 (2021).
4. Ko, E. et al. Is atelectasis related to the development of postoperative pneumonia? A retrospective single-center study. *BMC Anesthesiol* **23**, 77 (2023).
5. Newman, B., Krane, E. J., Gawande, R., Holmes, T. H. & Robinson, T. E. Chest CT in children: anesthesia and atelectasis. *Pediatr. Radiol.* **44**, 164–172 (2014).
6. Khan, A. et al. Severity of atelectasis during bronchoscopy: descriptions of a new grading system (ASSESS) and at-risk lung zones. *Diagnostics* **14**, 197 (2024).
7. McGuinness, S. & Parke, R. L. A new system for assessing atelectasis on chest X-ray after sternotomy for cardiac surgery. *Am. J. Respir. Crit. Care Med.* **183**, A1679 (2011).
8. Sargent, M. A., McEachern, A. M., Jamieson, D. H. & Kahwaji, R. Atelectasis on pediatric chest CT: comparison of sedation techniques. *Pediatr. Radiol.* **29**, 509–513 (1999).
9. Tenling, A., Hachenberg, T., Tydén, H., Wegenius, G. & Hedenstierna, G. Atelectasis and gas exchange after cardiac surgery. *Anesthesiology* **89**, 371–378 (1998).
10. Reber, A., Nylund, U. & Hedenstierna, G. Position and shape of the diaphragm: implications for atelectasis formation. *Anaesthesia* **53**, 1054–1061 (1998).
11. Warner, D. O., Warner, M. A. & Ritman, E. L. Atelectasis and chest wall shape during halothane anesthesia. *Anesthesiology* **85**, 49–59 (1996).
12. Pieters, A. L. P. et al. Structural lung disease and clinical phenotype in bronchiectasis patients: the EMBARC CT study. *Am. J. Respir. Crit. Care Med.* **210**, 87–96 (2024).
13. Rosenow, T. et al. PRAGMA-CF: a quantitative structural lung disease computed tomography outcome in young children with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **191**, 1158–1165 (2015).
14. Levine, B. E. & Johnson, R. P. Effects of atelectasis on pulmonary surfactant and quasi-static lung mechanics. *J. Appl. Physiol.* **20**, 859–864 (1965).
15. Noll, E. et al. Validation of MRI for volumetric quantification of atelectasis in the perioperative period: an experimental study in swine. *Front. Physiol.* **10**, 695 (2019).
16. David, M. et al. Analysis of atelectasis, ventilated, and hyperinflated lung during mechanical ventilation by dynamic CT. *Chest* **128**, 3757–3770 (2005).
17. Mizukami, Y., Takahashi, Y., Maki, R. & Adachi, H. Risk factors for atelectasis of the middle lobe after right upper lobectomy. *J. Thorac. Dis.* **13**, 5649–5657 (2021).
18. LOLA11 grand challenge. https://lola11.grand-challenge.org/ (2025).
19. Bragman, F. J. S. et al. Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior. *IEEE Trans. Med. Imaging* **36**, 1650–1663 (2017).
20. Imran, A. A. Z. et al. Fast and automatic segmentation of pulmonary lobes from chest CT using a progressive dense V-network. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **8**, 509–518 (2020).
21. Xie, W., Jacobs, C., Charbonnier, J.-P. & van Ginneken, B. Relational modeling for robust pulmonary lobe segmentation. *IEEE Trans. Med. Imaging* **39**, 2664–2675 (2020).
22. Peng, Y. & Zhang, J. Lung lobe segmentation based on multi-feature fusion and ensemble learning. *Int. J. Imaging Syst. Technol.* **33**, 2088–2099 (2023).
23. Pang, H. et al. A fully automatic segmentation pipeline of pulmonary lobes before and after lobectomy. *Comput. Biol. Med.* **147**, 105792 (2022).
24. Isensee, F. et al. nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024* 488–498 (Springer, 2024).
25. Malik, H. et al. Deep learning-based classification of chest diseases using X-rays, CT scans, and cough sound images. *Diagnostics* **13**, 2772 (2023).
26. Malik, H. & Anees, T. Multi-modal deep learning methods for classification of chest diseases. *PLoS ONE* **19**, e0296352 (2024).
27. Rajpurkar, P. et al. Deep learning for chest radiograph diagnosis. *PLoS Med.* **15**, e1002686 (2018).
28. Kamtam, D. N. et al. Does fixation of middle lobe/lingula during right upper lobectomy/left apical trisegmentectomy reduce postoperative torsion/atelectasis? *Ann. Thorac. Surg. Short Rep.* https://doi.org/10.1016/j.atssr.2026.03.017 (2026).
29. Tejani, A. S. et al. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol. Artif. Intell.* **6**, e240300 (2024).
30. Støverud, K. H. et al. AeroPath: an airway segmentation benchmark dataset with challenging pathology and baseline method. *PLoS ONE* **19**, e0311416 (2024).
31. Fedorov, A. et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **30**, 1323–1341 (2012).
32. Amini, E. & Klein, R. Lung lobe segmentation and quantification dataset. https://zenodo.org/records/12690803 (2024).
33. Isensee, F. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
34. Doel, T. et al. Pulmonary lobe segmentation from CT images using fissureness, airways, vessels and multilevel B-splines. In *Proc. IEEE International Symposium on Biomedical Imaging (ISBI)* 1491–1494 (IEEE, 2012).
35. Çiçek, Ö. et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Proc. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, 424–432 (Springer, 2016).
36. Park, J. et al. Fully automated lung lobe segmentation with 3D U-Net. *J. Digit. Imaging* **33**, 221–230 (2020).
37. Lassen-Schmidt, B., Hering, A., Krass, S. & Meine, H. Automatic segmentation of the pulmonary lobes with a 3D u-net and optimized loss function. In Medical Imaging with Deep Learning (MIDL) (2020).
38. Ferreira, F. T. et al. End-to-end supervised lung lobe segmentation. In *Proc. 2018 International Joint Conference on Neural Networks (IJCNN)* 1–8 (IEEE, 2018).
39. Gerard, S. E. & Reinhardt, J. M. Pulmonary lobe segmentation using convolutional neural networks. In *Proc. IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 1207–1211 (IEEE, 2019).
40. Zheng, S. et al. A dual-attention V-network for pulmonary lobe segmentation. *IET Image Process* **15**, 1644–1654 (2021).
41. Zhang, Z. et al. Automatic segmentation of pulmonary lobes on low-dose CT using deep learning. *Ann. Transl. Med.* **9**, 291 (2021).

42. Bao, N. et al. Edge-enhancement cascaded network for lung lobe segmentation. *Front. Phys*. **11**, 1098756 (2023).
43. Nomura, G. R. O. et al. TriSwinUNETR lobe segmentation model for computing DIR-free CT-ventilation. *Front. Oncol.* **15**, 1475133 (2025).
44. van Rikxoort, E. M. et al. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE Trans. Med. Imaging* **29**, 1286–1296 (2010).
45. Lassen, B. et al. Automatic segmentation of pulmonary lobes based on fissures, vessels, and bronchi. *IEEE Trans. Med. Imaging* **32**, 210–222 (2013).

Acknowledgements

Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. This research used data/services provided by STARR (STAnford medicine Research data Repository), a clinical data warehouse containing patients' health information of the Stanford hospitals made possible by Stanford School of Medicine Research Office and the Stanford Clinical and Translational Science Award Number UL1TR003142 from the National Center for Advancing Translational Sciences. Data from the Stanford customized version of The Society of Thoracic Surgery Database was also used for creating this study cohort.

Author contributions

D.N.K.—Conceptualization, data curation, formal analysis, methodology, supervision, writing—original draft, writing—review and editing; G.M.F.—Conceptualization, methodology, writing—review and editing; N.L.—Conceptualization, methodology, writing—review and editing; L.L.T.—Conceptualization, methodology, writing—review and editing; N.S.L.—Conceptualization, methodology, writing—review and editing; I.A.E.—Conceptualization, methodology, writing—review and editing; D.Z.L.—Conceptualization, methodology, writing—review and editing; L.M.B.—Conceptualization, methodology, writing—review and editing; M.F.B.—Conceptualization, methodology, writing—review and editing; H.H.G.—Conceptualization, methodology, writing—review and editing; C.P.L.—Conceptualization, methodology, writing—review and editing; J.B.S.—Conceptualization, methodology, supervision, writing—original draft, Writing—review and editing. All authors read and approved of the final manuscript.

Competing interests

Joseph Shrager: Consulting - Becton Dickinson; Lungpacer; Leah M. Backhus, advisory panel member with Johnson and Johnson, AstraZeneca, Genentech/Roche, and Bristol Myers Squibb; Natalie S. Lui, consulting with Intuitive Surgical Inc. and Centese, grants from Intuitive Foundation; Other authors have nothing to declare. Curt Langlotz - This research is supported in part by the Medical Imaging and Data Resource Center; grant (75N92020D00021) from National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health and through The Advanced Research Projects Agency for Health (ARPA-H). Dr. Langlotz has received consulting fees from Sixth Street; honoraria for lectures and support for travel from Singapore Ministry of Health, McKinsey, and Philips; patent pending with GE HealthCare; former president of RSNA; on board of directors and shareholder of Bunkerhill Health (since March 31, 2019); on board of directors Sirona Medical (since February 2025) option holder of Whiterabbit.ai (since October 1, 2017); advisor and option holder of Galileo CDS (since May 1, 2019); advisor and option holder of Sirona Medical (since July 6, 2020); advisor and option holder of ADRA.ai (since September 17, 2020); advisor and option holder of Cognita (since November 13, 2024); Advisor and shareholder, TurboRadiology (since January 17, 2025); and gifts to institution, department, and/or research center from BunkerHill Health, Carestream, CARPL.ai, Clairity, GE HealthCare, Google Cloud, IBM, Kheiron, Lambda, Lunit, Microsoft, Nightingale Open Science, Philips, Siemens Healthineers, Stability.ai, Subtle Medical, VinBrain, Visiana, Whiterabbit.ai, Lowenstein Foundation, and Gordon and Betty Moore Foundation. The other authors do not have a competing interest.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work, the author(s) utilized ChatGPT to assist with rephrasing and refining the writing in the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-026-02683-6.

Correspondence and requests for materials should be addressed to Devanish N. Kamtam.