RESEARCH PAPER

Quantifying emphysema in lung screening computed tomography with robust automated lobe segmentation

Thomas Z. Li [a,b,*] **Ho Hin Lee,** [a] **Kaiwen Xu,** [c] **Riqiang Gao,** [c] **Benoit M. Dawant,** [a, c,d,e] **Fabien Maldonado,** [f] **Kim L. Sandler,** [e] **and Bennett A. Landman** [a,c,d,e]

[a]Vanderbilt University, Department of Biomedical Engineering, Nashville, Tennessee, United States
[b]Vanderbilt University, School of Medicine, Nashville, Tennessee, United States
[c]Vanderbilt University, Department of Computer Science, Nashville, Tennessee, United States
[d]Vanderbilt University, Department of Electrical and Computer Engineering, Nashville, Tennessee, United States
[e]Vanderbilt University Medical Center, Department of Radiology and Radiological Sciences, Nashville, Tennessee, United States
[f]Vanderbilt University Medical Center, Department of Medicine, Nashville, Tennessee, United States

ABSTRACT. Purpose: Anatomy-based quantification of emphysema in a lung screening cohort has the potential to improve lung cancer risk stratification and risk communication. Segmenting lung lobes is an essential step in this analysis, but leading lobe segmentation algorithms have not been validated for lung screening computed tomography (CT).

Approach: In this work, we develop an automated approach to lobar emphysema quantification and study its association with lung cancer incidence. We combine self-supervised training with level set regularization and finetuning with radiologist annotations on three datasets to develop a lobe segmentation algorithm that is robust for lung screening CT. Using this algorithm, we extract quantitative CT measures for a cohort ($n = 1189$) from the National Lung Screening Trial and analyze the multivariate association with lung cancer incidence.

Results: Our lobe segmentation approach achieved an external validation Dice of 0.93, significantly outperforming a leading algorithm at 0.90 (p < 0.01). The percentage of low attenuation volume in the right upper lobe was associated with increased lung cancer incidence (odds ratio: 1.97; 95% CI: [1.06, 3.66]) independent of $PLCO_{m2012}$ risk factors and diagnosis of whole lung emphysema. Quantitative lobar emphysema improved the goodness-of-fit to lung cancer incidence ($\chi^2 = 7.48$, $p = 0.02$).

Conclusions: We are the first to develop and validate an automated lobe segmentation algorithm that is robust to smoking-related pathology. We discover a quantitative risk factor, lending further evidence that regional emphysema is independently associated with increased lung cancer incidence. The algorithm is provided at https://github.com/MASILab/EmphysemaSeg.

© 2023 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.10.4.044002]

Keywords: Lobar emphysema; lung screening; lung cancer risk; pulmonary lobe segmentation; level set method

Paper 22353GRR received Dec. 16, 2022; revised Jun. 14, 2023; accepted Jun. 21, 2023; published Jul. 18, 2023.

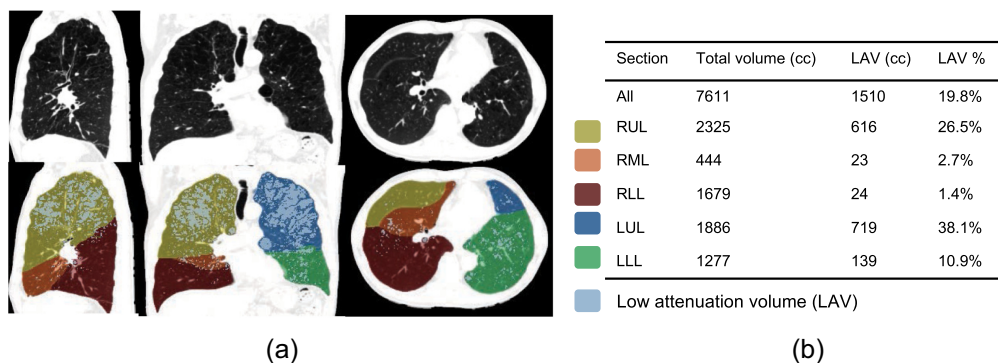*Address all correspondence to Thomas Z. Li, thomas.z.li@vanderbilt.edu

1 Introduction

Annual lung cancer screening with low dose computed tomography (CT) is the standard-of-care for individuals with a substantial smoking history,[1] and two of the goals include (1) estimating lung cancer risk early and (2) promoting health behavior change. Within the first goal, identifying novel risk factors of lung cancer works toward an individualized approach to lung cancer risk and further elucidates radiologic manifestations that are correlated with lung cancer development.[2–6] As a related second goal, annual lung screenings have served as encounters to communicate lung cancer risk and promote health behavior change such as smoking cessation. As a common abnormality seen on lung screening CT that correlates with tobacco use and the development of lung cancer, emphysema may be a useful biomarker for both goals.

Emphysema is inflammation and destruction of lung parenchyma that is primarily detected through CT imaging in patients with chronic obstructive pulmonary disease (COPD).[7] Emphysema has been well established as an independent risk factor for primary lung cancer. Many CT studies have linked qualitative emphysema assessment to increased lung cancer risk,[8–10] although these studies are subject to inter-reader variability. An automated quantitative approach to assessing emphysema avoids inter-reader variability and scales well with larger cohorts. Quantitative emphysema, also known as low attenuation volume (LAV), is approximated as intensity below −950 Hounsfield Units (HU). With this definition, the most up-to-date pooled meta-analysis[11] of studies from the United States,[12–16] Norway,[17,18] Japan,[19,20] and Spain[21] supports the independent association of quantitative emphysema severity with a higher risk of lung cancer.

However, recent literature has suggested that increased lung cancer risk is stratified by the location of emphysema. Bae et al.[22] found that upper lobe LAV and the ratio of normal attenuation volume and LAV were significant predictors of lung cancer development in the same lobe. These results highlight the need for emphysema measures across different pulmonary lobes in assessing lung cancer risk. In the aforementioned meta-analysis,[11] a pooled analysis of three studies[9,17,21] discovered that the association with lung cancer held for centrilobular but not paraseptal emphysema. The former is a subtype of emphysema that predominantly affects central and upper distributions of the lung, whereas the latter is distributed peripherally, adjacent to pleural surfaces. Together, these findings argue for the need to anatomically quantify emphysema and further investigate the effect of lobar emphysema on lung cancer risk.

Regarding the second goal, studies have shown that positive change in smoking habits have correlated with increased adherence to lung screening as well as the presence of an abnormal CT finding.[23–25] These studies support the importance of communicating lung screening results and suggest that integrated smoking cessation counseling may improve smoking abstinence. To this end, allowing patients to visualize emphysema in their lungs has the potential to improve communication and enhance behavior counseling in a shared decision-making setting. Clinical trials

Section	Total volume (cc)	LAV (cc)	LAV %
All	7611	1510	19.8%
RUL	2325	616	26.5%
RML	444	23	2.7%
RLL	1679	24	1.4%
LUL	1886	719	38.1%
LLL	1277	139	10.9%
Low attenuation volume (LAV)			

(a)　　　　　　　　　　　　　　(b)

Fig. 1 Lobe segmentation helps localize emphysema in across different lobes of the lung field. (a) Segmentation of lobes and LAV for a chest CT with high emphysema involvement. Emphysema is approximated as LAV below −950 HU. (b) Segmentation enables the measurement of percent emphysema involvement across pulmonary lobes. This information aids personalized radiologic evaluation and clinical management as well as disease characterization at scale for population-level research.

have been proposed to research this model of care,[26,27] but an automated tool for producing such visualizations has not been made available (Fig. 1).

In this study, we develop and validate a lobe segmentation algorithm that is robust to smoking-related pathology. We leverage this algorithm to quantify emphysema in a large lung screening cohort. We investigate the multivariate association of lobar LAV with lung cancer incidence after adjusting for an inclusive set of risk factors, including COPD. Finally, we ask if lobar LAV remains significantly associated with lung cancer risk independent of radiologist diagnosis of emphysema and whole lung LAV.

2 Datasets

This study involves deidentified human subjects and was supervised under Institutional Review Board #181279 titled "SPORE Pilot Project: Machine Learning for Prognosis Assessment" at Vanderbilt University. Informed consent was waived.

Vanderbilt lung screening program (VLSP)[28] is an on-going lung cancer screening program at Vanderbilt University Medical Center. Radiologist-annotated labels for the pulmonary lobes were not available for the VLSP. As a lung screening cohort, VLSP subjects have a 20 pack-year smoking history at minimum and have smoked within 15 years of scan acquisition. Across 887 subjects in the program, we collected 1490 chest CT scans (Table 1) that passed basic quality control, ensuring no artifact occluding the lung fields, proper field of view, slice contiguity, and realistic physical dimensions.[29]

TotalSegmentator (TotalSeg)[30] is a publicly available dataset of clinically collected CTs sampled from the University Hospital Basel, Switzerland, containing images of various protocols, slice thicknesses, resolutions, and reconstruction kernels. One hundred and four anatomical structures, including the pulmonary lobes, were annotated with the supervision of a board-certified radiologist. We selected images on which at least part of the lung field was visible and cropped them based on the boundaries of the lobe annotations, resulting in 660 unique CTs.

Lung nodule analysis (LUNA16).[31] Data in the LUNA16 grand challenge were collected from LIDC-IDRI, a publicly available reference database of diagnostic and lung screening chest CTs. Tang et al.[32] publicly released manual annotations for 50 examples in this dataset. Our team manually reviewed this dataset and identified a subset of 47 annotated examples that were appropriate for the finetuning step. Reasons for exclusion included the presence of an artifact occluding the lung fields and incorrect annotations.

National Lung Screening Trial (NLST)[1] is a publicly available lung screening dataset. Its subjects have substantial smoking history and underwent annual low-dose chest CTs. Because corresponding demographics, risk factors, smoking history, and past medical diagnoses are also available, these risk factors were included to study the adjusted association between quantitative lobar emphysema and lung cancer incidence. For this dataset, we sampled a cohort of 578 biopsy-confirmed lung cancer cases and 611 controls randomly sampled from subjects who did not develop lung cancer 2 years after their latest scans. Quantitative emphysema measures were extracted from a soft kernel CT from the first screening session of each subject in this cohort. Characteristics of the cohort are summarized in Table 2.

Table 1 Characteristics of datasets used to develop the lobe segmentation model.

Characteristics	VLSP	Total-segmentator	LUNA16
Population	Lung screening	Clinical routine	Lung screening
Number of subjects	887	660[b]	47
Number of scans	1490	660[b]	47
Annotations	Pseudo-labels[a]	Radiologist labels	Radiologist labels

[a]Pseudo-labels were generated using methods described in Sec. 3.1.
[b]Scans with pulmonary lobe annotations available.

Table 2 Characteristics of a balanced sample of lung cancer cases and controls from NLST.

	Lung cancer cases	Controls
Number of subjects/scans	578	611
Age	64 ± 5	62 ± 5
Race		
White	541 (93.6%)	575 (94.1%)
Black	24 (4.2%)	14 (2.3%)
Hispanic or Latino	7 (1.2%)	10 (1.6%)
Other[a]	6 (1%)	12 (2%)
Education		
Less than high school	54 (9.3%)	35 (5.7%)
High school or GED	152 (26.3%)	145 (23.7%)
Post high school training, excluding college	84 (14.5%)	95 (15.5%)
Associate's degree	128 (22.1%)	153 (25.0%)
Bachelor's degree	86 (14.9%)	102 (16.7%)
Graduate degree	63 (10.9%)	70 (11.5%)
Other[b]	11 (1.9%)	11 (1.8%)
BMI (kg/m^2)	26.89 ± 4.70	27.85 ± 4.80
COPD present	66 (11.4%)	34 (5.5%)
Personal cancer history	35 6.0%	27 4.4%
Family lung cancer history	149 (25.8%)	144 (23.6%)
Smoking status		
Former	253 (43.8%)	336 (55.0%)
Current	325 (56.2%)	275 (45.0%)
Smoking quit years	2.71 ± 4.27	3.86 ± 4.86
Smoking pack-years	65.12 ± 27.42	57.20 ± 25.98
Radiologic emphysema	318 (55.0%)	266 (43.5%)
Lung volume (cc)	5774 ± 1307	5637 ± 1294
LAV %		
Total	8.0 [1.7, 12.2]	6.9 [1.3, 9.6]
LUL	7.8 [1.7, 9.8]	7.2 [1.3, 8.7]
LLL	5.7 [1.0, 6.2]	5.6 [0.7, 5.9]
RUL	12.8 [1.4, 15.2]	9.8 [1.0, 10.0]
RML	7.1 [1.5, 9.3]	6.9 [1.5, 9.7]
RLL	4.8 [0.7, 5.5]	4.2 [0.6, 5.1]

Continuous variables are given in mean ± standard deviation or mean [First quartile, third quartile]. BMI = body mass index. COPD = chronic obstructive pulmonary disease. LAV % = low attenuation volume percentage.
[a]Asian, American Indian, Alaskan Native, Native Hawaiian, Pacific Islander, missing value, or decline to answer.
[b]Missing value or decline to answer.

3 Methods

3.1 Lobe Segmentation

Characterizing emphysema across a large lung screening cohort requires automatic lobe segmentation that is robust to smoking-related pathology. The leading algorithm from Hofmanninger et al.[33] has been validated on diverse lung pathology but exhibits substantial artifact at the pulmonary fissures (Fig. 2) that are not consistent with realistic 3D anatomy. To overcome this challenge, we employed a volumetric level set method[34] (LSM) to infer smoothly contoured and artifact-free borders [Fig. 3(b)] from the leading algorithm. LSM-evolved segmentations were acquired for each scan in VLSP and used as pseudo-labels to train a 3D U-Net from random weights [Fig. 3(c)]. The resulting model was then finetuned with near ground-truth annotations from TotalSeg [Fig. 3(d)]. The final model was quantitatively validated on an external screening CT dataset and qualitatively reviewed for a sample hold out set of VLSP [Fig. 3(e)].
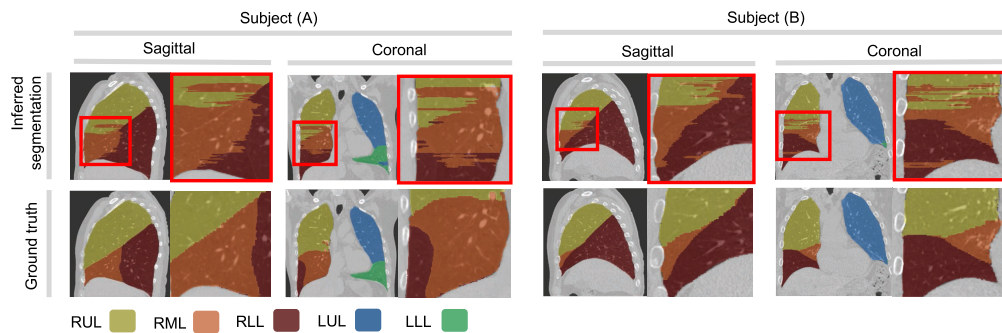
Fig. 2 State-of-the-art lobe segmentation algorithms fail to capture realistic lobe fissures in 3D in lung screening CT. The leading lobe segmentation model from Hofmanninger et al.[33] applied to two examples. An irregular artifact is noticeable along the borders of the segmentation. In addition, the right middle lobe (orange) and right lower lobe (red) are inaccurately segmented.
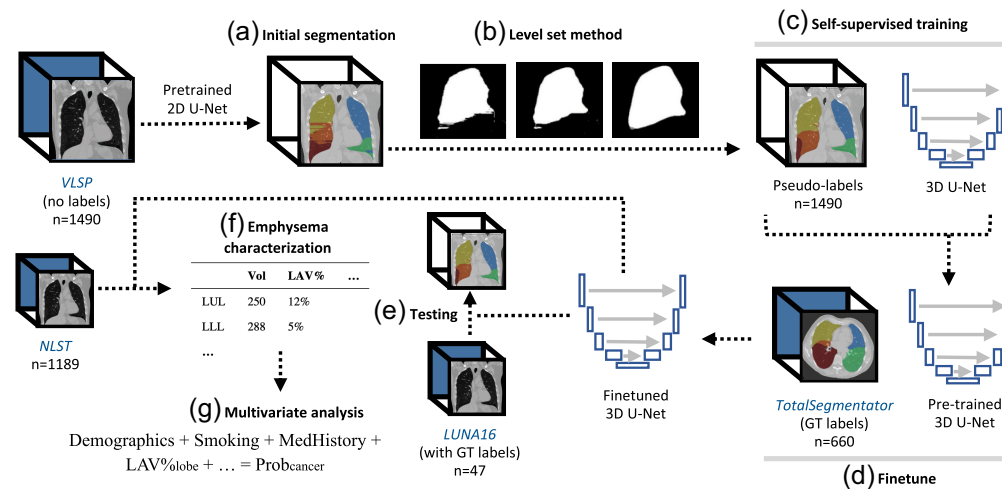
Fig. 3 Overview of methods. (a) The 2D U-Net based segmentation algorithm from Hofmanninger et al.[33] is applied to infer initial segmentations for the VLSP cohort. Pulmonary lobe annotations are absent for this cohort. (b) Initial lobe segmentations are iteratively evolved using a volumetric level set method (LSM). (c) The evolved segmentations are used as pseudo-labels for VLSP in the self-supervised training of our 3D U-Net. (d) The model is further finetuned with 660 annotated chest CTs from the TotalSegmentator dataset. (e) An external validation set from LUNA16 is used to test the model's performance on lung screening CTs. (f) The final model is leveraged to anatomically quantify emphysema severity for the NLST cohort. (g) Multivariate logistic regression is applied to study the association between emphysema measures and lung cancer incidence.

3.1.1 *Level set method*

The LSM provides a framework to efficiently compute an evolving surface as discrete points by embedding it as a level set of a higher-dimensional function. We leverage a classic LSM formulation for image segmentation that has been well documented in the literature.[35–38] In this formulation, a lobe segmentation is computed from the zero level set, S, of a time-varying function, $\phi : \mathbb{R}^4 \to \mathbb{R}$ as

$$S = \{x(t) | \phi(x(t), t) = 0\}.$$

The zero level set is therefore the set of discrete points at t where $\phi = 0$, and this set is assumed to approximate an isosurface that represents the segmentation evolved after t timesteps. ϕ is a function that models a 4D hypersurface by mapping a position at t to its signed distance to the zero level set. We approximate the initial value of ϕ at $t = 0$ as a signed distance transform of the initial segmentation of a single lobe. The distance transform of each voxel inside the segmentation is computed as the negative Euclidean distance to the nearest background voxel. This results in a distance map in which voxels near the center of the segmentation were more negative than those near the edge. The value of background voxels remains zero. The implicit level set model is then iteratively evolved according to handcrafted force functions designed to drive the model toward a desired deformation. During evolution, the evolution speed is derived from an initial ϕ at $t = 0$ by differentiating Eq. (1) with respect to t as

$$\phi(x(t), t) = 0,$$
$$\frac{\partial \phi(x(t), t)}{\partial t} = 0, \tag{1}$$

$$\frac{\partial \phi}{\partial t} = -\nabla \phi \frac{\partial x(t)}{\partial t}, \tag{2}$$

where $\frac{\partial x(t)}{\partial t}$ is expressed as the evolution speed, F, in the surface's outward normal direction such that $\frac{\partial x(t)}{\partial t} = F \frac{\nabla \phi}{\|\nabla \phi\|}$. In imaging applications, F is a function of local properties of the image and level surface. Although there are many choices of local properties to operate on, we found the combination of a boundary attachment term and a smoothing term to be optimal for our purpose, which is given as

$$F = \alpha \nabla \frac{1}{1 + \|G(b, \sigma)\|} + \beta \operatorname{div}\left(\frac{\nabla \phi}{\|\nabla \phi\|}\right), \tag{3}$$

where $G(b, \sigma)$ is the binary segmentation map, b, filtered with a Gaussian kernel of standard deviation σ. The first term encourages the zero level surface to remain close to the initial segmentation boundary within a range directly proportional to σ. This is a desirable property because our goal with the LSM is to correct local artifacts at the border without substantially changing the general position or volume of the initial segmentations. A larger σ allows the zero level surface to evolve further from the initial segmentation boundary, and vice versa. The second term is a measure of the surface's local mean curvature. It encourages deformations along the normal direction in regions of high curvature, thereby smoothing the surface. Positive constants α and β govern the weight of the boundary attachment term and the smoothing term, respectively. Solving the partial differential equation in Eq. (2) with the finite difference method results in the zero level surface at time t

$$\phi(x(t), t + \Delta t) = \phi(x(t), t) + \Delta t F \|\nabla \phi\|. \tag{4}$$

A separate model, $\phi(x(0), 0)_l$, is initialized for each lobe volume, V_l, which is segmented using the algorithm from Hofmanninger et al.[33] Each model is evolved according to Eq. (4) for a fixed Δt and fixed number of iterations, k. The final zero level surfaces are computed as $\phi(x(k\Delta t), k\Delta t)_l = 0$, and the volume enclosed by this isosurface, denoted by $\hat{V}_l$, is computed as $\phi(x(k\Delta t), k\Delta t)_l \leq 0$. The five volumes, $\{\hat{V}_1, \ldots, \hat{V}_5\}$, acquired in this way, are merged according to the following scheme: (1) a global lung field is computed as the union of all $\hat{V}_l$.

(2) Voxels in the lung field with more than one label or no label are assigned to $\hat{V}_l$, where l corresponds to the smallest valued ϕ_l. (3) The largest connected component of each $\hat{V}_l$ is found.

Herein we briefly describe the specific settings and hyperparameters that were empirically found to be optimal for the practical implementation of the LSM on the VLSP dataset. Resizing initial volumes to $512 \times 512 \times 320$ results in a LSM that is efficient in runtime and memory without a noticeable tradeoff in resolution. The parameters of F consist of a Gaussian kernel with $\sigma = 1.5$ as well as $\alpha = 1$ and $\beta = 1$.

3.1.2 *Self-supervised pretraining*

As shown in previous work,[33] achieving accurate, reliable, and robust lung segmentation in clinically collected medical images is more dependent on data diversity and quality than model choice. As such we refrained from developing a bespoke model architecture and used a generic state-of-the-art 3D U-Net.[39] The model operates on $96 \times 96 \times 96$ volumetric patches, and we leverage the LSM-corrected lobe segmentations as regularized pseudo-labels for self-supervised training. An 80% to 20% training-validation split of VLSP was used to train the 3D U-Net from random weights. During training, patches centered on a voxel in the lung field were randomly extracted and augmented with random intensity shifts and random affine transformations. Using the average of Dice loss and cross entropy loss as the training and validation criterion, the model was trained until the running average of the validation Dice decreased from the maximum observed by >10%.

3.2 Finetune Training and Validation

The model was finetuned with 660 CT scans and corresponding labels from TotalSegmentator using the same settings for data augmentation and loss criterion. Finetune training continued until the running average of the validation Dice decreased 10% from its peak. The finetuned model was tested using an external validation set consisting of 47 lung screening chest CTs from LUNA16 with corresponding radiologist-generated labels. During inference, our algorithm found connected components and set missing labels to their nearest neighbors as post-processing of U-Net outputs. For comparison, we computed the Dice score of original and LSM-corrected segmentations from Hofmanninger et al.[33] We report Dice scores for each pulmonary lobe as well as the average across the validation set (Fig. 4).

3.2.1 *Sensitivity analysis*

To access the segmentation of our proposed approach, the five finetuned models were tested on their corresponding hold-out folds. For comparison, we computed the Dice score of original and
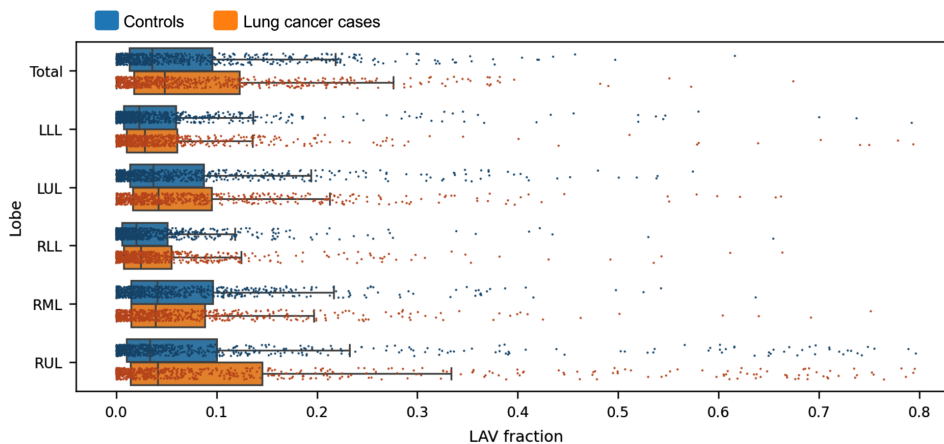
Fig. 4 LAV fraction across pulmonary lobes for lung cancer cases and controls without lung cancer. Greater variance in LAV fraction was observed in the RUL compared with other lobes.

Table 3 Characteristics of VLSP sample used for sensitivity analysis.

Lung cancer	COPD	Sex	Count
Present	Present	M	11
	Present	F	6
	Absent	M	11
	Absent	F	10
Absent	Present	M	35
	Present	F	37
	Absent	M	37
	Absent	F	39
Total			186

LSM-corrected segmentations from Hofmanninger et al.[33] We report the average Dice for each pulmonary lobe as well as the average across all examples from the finetuned dataset.

To access the reliability and the robustness of our proposed approach in the context of a lung screening population, we conducted a qualitative sensitivity analysis. Lobe segmentations for a sample from VLSP were inferred using one of our finetuned models. This sample (Table 3) was curated using a random selection stratified on COPD status (present or absent) and sex (M or F). In addition, the sample included all VLSP subjects with biopsy-diagnosed lung cancer. All scans for each subject in the sample were included in the analysis. The inferred lobe segmentations were scored 1 through 5 using the following criteria.

1. All lobes appear to be inaccurately segmented and/or segmentation artifact prevents assessment.
2. 2-3 lobes appear to be inaccurately segmented and/or there is severe segmentation artifact present.
3. 1-2 lobe segmentations appear to be <80% accurate and/or there is moderate segmentation artifact present.
4. All lobe segmentations appear to be >80% accurate and/or there is minor segmentation artifact present.
5. All lobes appear to be accurately segmented and there is no noticeable segmentation artifact present.

Scores from scans of the same patient were averaged into a single subject-level score. A two-tailed Welch's t-test was used to determine if scores were significantly different between male and female, lung cancer present and absent, and COPD present and absent populations.

3.3 Emphysema Characterization

We investigated the association between lobar emphysema and lung cancer in a lung screening cohort sampled from NLST. Emphysema was approximated as the volume below the standard low attenuation threshold of −950 HU. The LAV was intersected with labels inferred from our lobe segmentation model to quantify the LAV percentage in each lobe. Whole lung volume was computed by multiplying the voxel count with the voxel dimensions.

3.4 Statistical Analysis

We adopted the practice of Labaki et al.[16] in adjusting for covariates from the $PLCO_{m2012}$[6] lung cancer risk model, which were found to be significant predictors of developing lung cancer. These include age, sex, body mass index (BMI), COPD status, past cancer history, past lung cancer family history, smoking status, smoking quit time, and smoking pack years. We inferred

missing values of BMI and/or COPD status for eight subjects using multivariate imputation. Continuous covariates were min-max normalized with the following equation:

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}},$$

where x_i is the covariate of subject i, $x_{\min}$ is the minimum value of covariate x observed in the population, and $x_{\max}$ is the maximum value of covariate x observed in the population. A logistic regression model with lobar LAV percentage, whole lung volume, and PLCO$_{m2012}$ covariates was fit to estimate lung cancer risk. To further investigate if lobar LAV contributes to lung cancer risk prediction independently of whole lung emphysema, we fit two additional models that included radiologic emphysema and total LAV. The former is derived from abnormalities detected by radiologists as part of the NLST archive, whereas the latter was computed as the aggregate of lobar LAV percentage. Adjusted odds ratios and their p-values were computed for each quantitative CT measure (Table 4). The added value of lobar LAV was assessed by comparing the goodness-of-fit of PLCO$_{m2012}$ only and with the addition of lobar LAV, lobar LAV + radiologic emphysema, or lobar LAV + total LAV. The likelihood-ratio test was used to determine a significant change in goodness-of-fit, and the χ^2 statistic and p-value from this test are reported (Table 5).

Table 4 Adjusted odds ratios for lobar LAV with respect to lung cancer risk.

	Lobar LAV		+ Radiologic emphysema		+ Total LAV	
	Odds ratio[a] (95% CI)	p	Odds ratio[a] (95% CI)	p	Odds ratio[a] (95% CI)	p
Radiologic emphysema	N/A		1.21 (0.934, 1.56)	0.15	N/A	
Lung volume	1.46 (0.443, 4.81)	0.53	1.43 (0.433, 4.74)	0.56	1.20 (0.457, 3.14)	0.56
LAV %						
Total	N/A		N/A		0.078 (0.00, 276)	0.56
LUL	0.933 (0.252, 3.45)	0.92	0.810 (0.216, 3.04)	0.76	1.71 (0.155, 18.9)	0.66
LLL	0.525 (0.145, 1.90)	0.33	0.545 (0.151, 1.97)	0.36	1.09 (0.70, 17.1)	0.95
RUL	1.98 (1.07, 3.68)	0.03	1.89 (1.02, 3.52)	0.04	4.50 (0.273, 74.1)	0.29
RML	0.490 (0.070, 3.48)	0.48	0.537 (0.075, 3.85)	0.54	0.582 (0.075, 4.49)	0.61
RLL	2.71 (0.272, 26.9)	0.40	2.70 (0.268, 27.2)	0.40	4.89 (0.240, 99.8)	0.30

[a]Odds ratio is adjusted with PLCO$_{m2012}$ lung cancer covariates: age, race, education, BMI, COPD, past cancer history, past lung cancer family history, smoking status, smoking quit time, and smoking pack years.

Table 5 Added value of lobar LAV to goodness-of-fit.

	Likelihood-ratio test	
	χ^2	p
PLCOm2012	ref.	N/A
Lobar LAV	7.48	0.02
+ Radiologic emphysema	9.57	<0.01
+ Total LAV	7.82	0.02

Dice performance of lobe segmentation is reported as the mean from the LUNA16 test set. 95% confidence intervals for the mean Dice were computed from 100 bootstrap samples by sampling with replacement from the predictions of LUNA16. We used a two-tailed Wilcoxon signed rank test to determine if there was a significant difference in Dice performance between our approach and baselines with a significance level of $p < 0.05$.
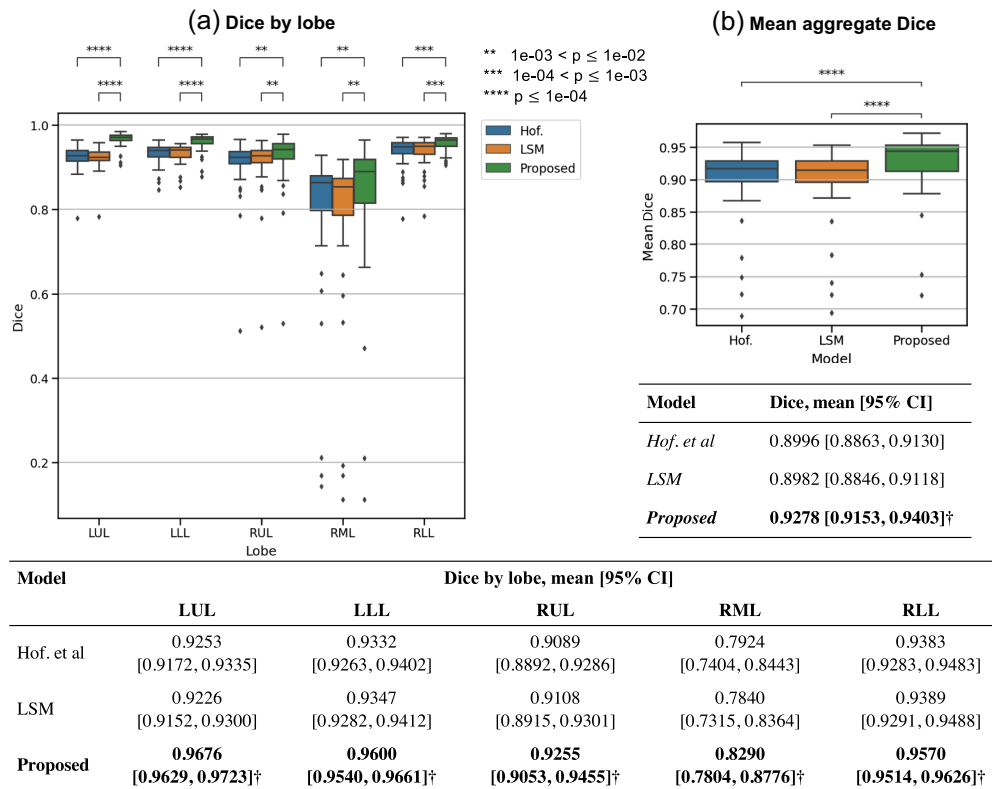
4 Results

4.1 Emphysema Characterization

4.1.1 *Quantitative CT measures*

The mean total LAV percentage in lung cancer cases (8.0 [1.7, 12.2]) was higher than that of controls (6.9 [1.3, 9.6]). The comparison remained true across all lobes, with the largest difference being observed in the RUL. Across all subjects, we observed a greater variance in the LAV percentage in the RUL compared with other lobes (Fig. 4). In addition, we observed a 140cc increase in mean whole lung volume of lung cancer cases compared with controls. These observations align with the increased prevalence of radiologic emphysema and COPD diagnosed in lung cancer cases (emphysema: 55%; COPD: 11.4%) compared with controls (emphysema: 44%; COPD: 5.5%).

4.1.2 *Correlation to lung cancer risk (Table 4)*

The LAV percentage in the RUL was found to be a significant predictor of lung cancer with an odds ratio of 1.97 (95% CI: [1.07, 3.68]). Its statistical significance held when controlling for

Model	Dice, mean [95% CI]
Hof. et al	0.8996 [0.8863, 0.9130]
LSM	0.8982 [0.8846, 0.9118]
Proposed	**0.9278 [0.9153, 0.9403]†**

Model	Dice by lobe, mean [95% CI]				
	LUL	LLL	RUL	RML	RLL
Hof. et al	0.9253 [0.9172, 0.9335]	0.9332 [0.9263, 0.9402]	0.9089 [0.8892, 0.9286]	0.7924 [0.7404, 0.8443]	0.9383 [0.9283, 0.9483]
LSM	0.9226 [0.9152, 0.9300]	0.9347 [0.9282, 0.9412]	0.9108 [0.8915, 0.9301]	0.7840 [0.7315, 0.8364]	0.9389 [0.9291, 0.9488]
Proposed	**0.9676 [0.9629, 0.9723]†**	**0.9600 [0.9540, 0.9661]†**	**0.9255 [0.9053, 0.9455]†**	**0.8290 [0.7804, 0.8776]†**	**0.9570 [0.9514, 0.9626]†**

Fig. 5 Mean Dice similarity coefficient across all cross validation folds on LUNA16. "Hof. et al." is the algorithm from Hofmanninger et al.[33] The LSM is evolved from Hof. et al using level set regularization and connected components. Our proposed model is trained using self-supervision and further fine-tuned on five cross validation folds with the combined LUNA16 and active learning dataset. † $p < 0.05$.

radiologic emphysema but not when controlling for the total LAV percentage. The odds ratio of the RLL LAV percentage was relatively high at 2.71(95% CI: [0.272, 26.9]) but was not found to be significantly different than 1 (Table 4). Compared with $PLCO_{m2012}$, the addition of lobar LAV significantly improved the goodness-of-fit ($\chi^2 = 7.48$; $p = 0.02$) according to the likelihood ratio test. The addition of radiologic emphysema ($\chi^2 = 9.57$; $p < 0.01$) and total LAV ($\chi^2 = 7.82$; $p = 0.02$) also significantly improved the goodness-of-fit (Table 5).

4.2 Lobe Segmentation

4.2.1 *Quantitative validation*

There was no significant difference in Dice performance between the baseline and LSM-processed results. In contrast, our proposed method (0.9278 Dice) significantly outperformed the baseline (0.8996 Dice) and LSM (0.8982 Dice). Across each pulmonary lobe, the baseline and LSM achieved a similar Dice performance, whereas our proposed method achieved a significantly higher average Dice. A two-tailed Wilcoxon signed rank test found these differences to be significant for $p < 0.05$. All three methods performed worse on the RML compared with the other lobes, but here the improvement from our proposed method was pronounced (Fig. 5).

4.2.2 *Qualitative sensitivity analysis*

No significant difference in the mean score between cancer status, COPD status, and sex were observed (Table 6). The mean scores for all clinical categories were above 4. Across all three categories, the score of 5 was the most frequent, reflecting the observation that these segmentations appeared to be accurate without any noticeable artifact. Qualitative performance was

Table 6 Qualitative scores by cancer, COPD, and sex.

	Mean score	Std	P
Cancer present	4.45	0.72	<0.05
Cancer absent	4.25	0.85	
COPD present	4.16	0.87	<0.05
COPD absent	4.34	0.80	
Male	4.03	0.86	<0.05
Female	4.51	0.73	
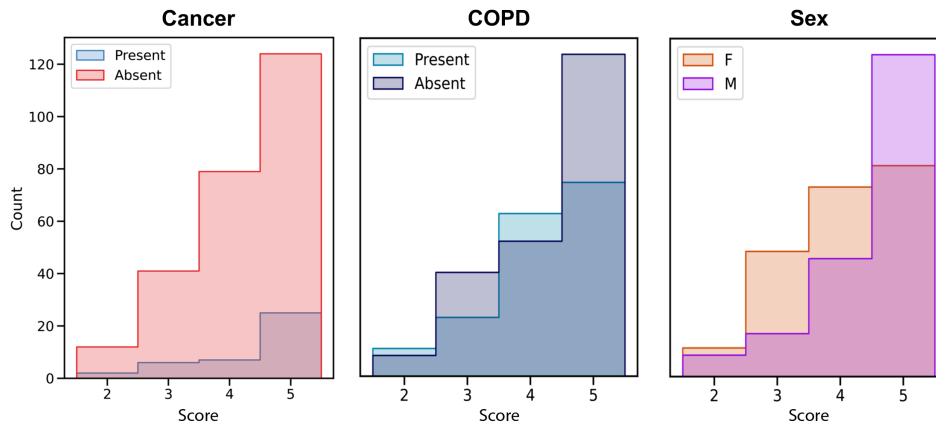
Fig. 6 Distributions of qualitative scores for a random sample of VLSP. The distributions of scores are heavily skewed toward a score of 4 and 5 regardless of lung cancer status, COPD status, and sex.

limited by a fewer number of 4s and 3s. These segmentations were ones that appeared less accurate and/or had some noticeable artifact, most commonly in subjects with challenging RML anatomy and obscure lobe fissures. A small number of scans received a score of 2, and no scans received a score of 1 (Fig. 6).

5 Discussion

### 5.1 Emphysema Characterization

We investigated lobar emphysema as an independent risk factor for lung cancer by applying an automated pipeline to quantify emphysema markers for a large lung screening cohort. Based on the quantitative percentage of LAV, we found higher emphysema involvement in the upper lobes, especially the RUL, compared with the lower lobes. These findings agree with previous studies of upper lobe-predominant emphysema being highly prevalent in lung screening cohorts.[9,16,40] The finding that LAV in the RUL independently contributes to lung cancer risk is novel. Our findings remain significant even when controlling for whole lung emphysema diagnosis, reinforcing the independent contribution of regional emphysema on lung cancer risk. Our results align with a previous study[22] that found upper lobe emphysema to be a significant predictor of lung cancer location, although the distinction between left and right upper lobes was not made in that study. Our results also agree with a pooled meta-analysis of three studies[9,17,21] showing that centrilobular emphysema, which presents with upper lobe predominance, is independently associated with increased lung cancer risk. We did not confirm whether centrilobular emphysema was driving the observed association in our cohort because diagnosing emphysema subtype requires a visual assessment.

The lack of significant association between the LUL LAV percentage and lung cancer in our cohort is surprising given that the upper lobes are thought to be similarly affected in emphysema. The large odds ratio of the RLL LAV percentage with a 0.395 p-value is also worth noting and suggests a distinction between left and right lungs that has not been studied in the context of emphysema and lung cancer. The right main bronchus is wider, shorter, and more vertical than the left. The emphysema quantification of the LUL combines the LUL and the lingula, which is anatomically like the RML. However, the left and right upper lungs are thought to be physiologically similar in ventilation and perfusion despite these anatomical differences. Evidence of distinction between left and right lungs is more likely specific to this study's cohort, and validation in other cohorts is needed for further investigation.

This section of work is limited in several aspects. First, Caucasians are overrepresented in both cases and controls of the NLST cohort. This may bias the results and limit their applicability to populations of greater racial diversity. Second, inaccurate segmentations may introduce noise into the quantitative emphysema measures, especially when quantifying LAV in the RML. We strived to minimize this noise by optimizing our model for lung screening CT and standardizing the analysis to soft kernel CTs. Third, large emphysematous bullae may confound emphysema quantification as these structures are known to compress surrounding lung tissue. This would nonlinearly influence LAV-based quantification.

5.2 Lobe Segmentation

The present work seeks to develop a lobe segmentation algorithm that appears realistic in 3D and is robust to smoking-related changes in the lung. To this end, we proposed a two-stage training strategy that combines self-supervised training on a lung screening dataset with level set regularization and finetuning on a clinical routine dataset with near ground truth labels. In general, the LSM was successful in smoothing local artifacts, such as those in subjects A and B in Fig. 7, but it preserved large scale defects such as those in subjects C. Although such defects introduced label noise into the first stage of training, the final model resolved both small- and large-scale artifacts. Testing on an external validation set confirmed that our algorithm outperformed a leading baseline method on lung screening subjects across all lung lobes. As the most variable in anatomical shape and position, the RML was unsurprisingly the most challenging to segment. Although superior to the leading method, our algorithm was only moderately accurate in segmenting the RML.
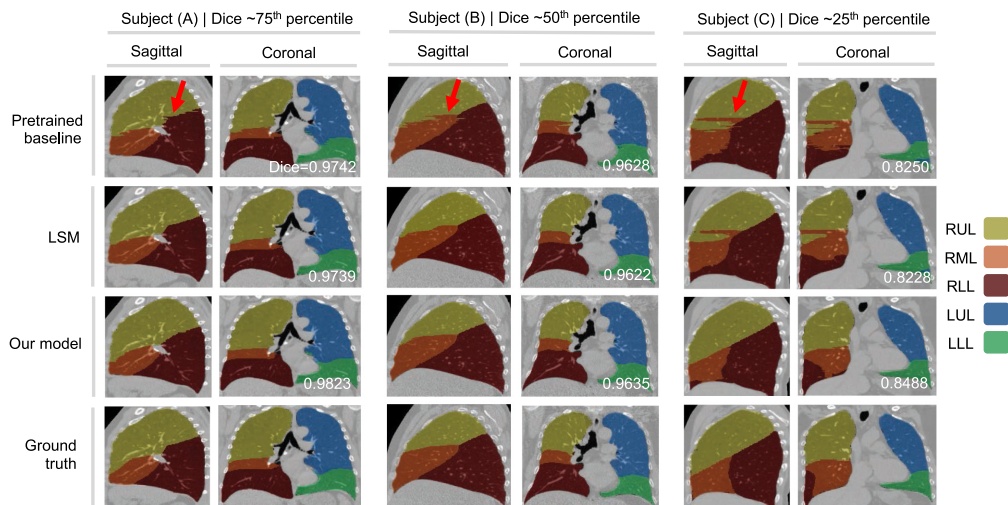
Fig. 7 Lobe segmentations and ground truth annotations for three subjects using three methods: a pretrained baseline model, the baseline model post-processed with the LSM, and our proposed model. Segmentations of subject A, B, and C come from the 75'th, 50'th, and 25'th percentiles of our proposed model's Dice performance. The LSM resolves small-scale border artifacts and achieves smoothly contoured fissures, but large-scale defects such as those seen in subject C's RML remain unresolved. Our proposed method effectively resolves large-scale defects and segments the RML more accurately than the baseline method.

Our algorithm was not free of errors, as suggested by a mix of scores in our sensitivity analysis. Common failure modes included inaccurate segmentation of challenging RML anatomy and presence of visually apparent artifacts (Fig. 7). The scores were calibrated such that a 3 or above would be acceptable for use in most visualization applications. Because the overwhelming majority of samples scored in the acceptable range, this analysis supported the clinical utility of our algorithm. Because our algorithm was primarily developed and validated on lung screening subjects who all have substantial smoking history, its performance on pulmonary pathologies seen in the clinical routine is unknown.

6 Conclusion

We developed an automated pipeline for robust quantification of emphysema and used it to investigate the association between lobar emphysema and lung cancer. We employed self-supervised training with level set regularization and ground truth finetuning to maximize our model's performance on smoking-related pathology and minimize its susceptibility to producing border artifacts. As a result, our lobe segmentation algorithm is more accurate on lung screening CT compared with a leading baseline and is optimally suited to quantify lobar emphysema. The algorithm is made publicly available at https://github.com/MASILab/EmphysemaSeg. We quantified emphysema for a large lung screening cohort and are the first to find that a high LAV in the RUL is an independent risk factor for lung cancer.

Disclosures

No conflicts of interests are reported.

Data Availability

The data presented in this article from the NLST are available upon request at https://cdas.cancer .gov/learn/nlst/images/. The data from LUNA16 are available at https://luna16.grand-challenge.org/ Home/, and lobe annotations are released at https://github.com/deep-voxel/automatic_pulmonary_ lobe_segmentation_using_deep_learning. The data from TotalSegmentator are publicly available at https://github.com/wasserth/TotalSegmentator. The data from VLSP used in this article are not publicly available due to institutional and privacy restrictions.

References

1. "Reduced lung-cancer mortality with low-dose computed tomographic screening," *N. Engl. J. Med.* **365**, 395–409 (2011).
2. P. P. Massion and R.C. Walker, "Indeterminate pulmonary nodules: risk for having or for developing lung cancer?" *Cancer Prevent. Res.* **7**, 1173–1178 (2014).
3. R. Clay et al., "Computer Aided Nodule Analysis and Risk Yield (CANARY) characterization of adeno-carcinoma: radiologic biopsy, risk stratification and future directions," *Transl. Lung Cancer Res.* **7**, 313–326 (2018).
4. R. Gao et al., "Cancer risk estimation combining lung screening CT with clinical data elements," *Radiol. Artif. Intell.* **3**, e210032 (2021).
5. M. N. Kammer et al., "Integrated biomarkers for the management of indeterminate pulmonary nodules" (2021).
6. M. C. Tammemägi et al., "Selection criteria for lung-cancer screening," *N. Engl. J. Med.* **368**, 728–736 (2013).
7. S. A. Christenson et al., "Chronic obstructive pulmonary disease," *Lancet* **399**, 2227–2242 (2022).
8. P. C. Yong et al., "The effect of radiographic emphysema in assessing lung cancer risk," *Thorax* **74**, 858–864 (2019).
9. J. González et al., "Emphysema phenotypes and lung cancer risk," *PLoS One* **14**, e0219187 (2019).
10. C. I. Henschke et al., "CT screening for lung cancer: importance of emphysema for never smokers and smokers," *Lung Cancer* **88**, 42–47 (2015).
11. X. Yang et al., "Association between Chest CT–defined emphysema and lung cancer: a systematic review and meta-analysis," *Radiology* **304**, 322–330 (2022).
12. F. Maldonado et al., "Are airflow obstruction and radiographic evidence of emphysema risk factors for lung cancer? A nested case-control study using quantitative emphysema analysis," *Chest* **138**, 1295–1302 (2010).
13. D. S. Gierada et al., "Quantitative CT assessment of emphysema and airways in relation to lung cancer risk," *Radiology* **261**, 950–959 (2011).
14. L. L. Carr et al., "Features of COPD as predictors of lung cancer," *Chest* **153**, 1326–1335 (2018).
15. A. G. Schwartz et al., "Risk of lung cancer associated with COPD phenotype based on quantitative image analysis," *Cancer Epidemiol. Biomarkers Prev.* **25**, 1341 (2016).
16. W. W. Labaki et al., "Quantitative emphysema on low-dose CT imaging of the chest and risk of lung cancer and airflow obstruction: an analysis of the National Lung Screening Trial," *Chest* **159**, 1812–1820 (2021).
17. A. A. Gagnat et al., "Incidence of non-pulmonary cancer and lung cancer by amount of emphysema and airway wall thickness: a community-based cohort," *Eur. Respir. J.* **49**, 1601162 (2017).
18. G. R. Husebø et al., "Risk factors for lung cancer in COPD - results from the Bergen COPD cohort study," *Respir. Med.* **152**, 81–88 (2019).
19. M. Nishio, T. Kubo, and K. Togashi, "Estimation of lung cancer risk using homology-based emphysema quantification in patients with lung nodules," *PLoS One* **14**, e0210720 (2019).
20. S. Chubachi et al., "Radiologic features of precancerous areas of the lungs in chronic obstructive pulmonary disease," *Int. J. Chron. Obstruct. Pulmon. Dis.* **12**, 1613–1624 (2017).
21. C. Mouronte-Roibás et al., "Influence of the type of emphysema in the relationship between COPD and lung cancer," *Int. J. Chron. Obstruct. Pulmon. Dis.* **13**, 3563 (2018).
22. K. Bae et al., "Severity of pulmonary emphysema and lung cancer: analysis using quantitative lobar emphysema scoring," *Medicine* **95**, e5494 (2016).
23. C. M. van der Aalst et al., "The impact of a lung cancer computed tomography screening result on smoking abstinence," *Eur. Respir. J.* **37**, 1466–1473 (2011).
24. C. G. Slatore et al., "Smoking behaviors among patients receiving computed tomography for lung cancer screening. Systematic review in support of the U.S. preventive services task force," *Ann. Am. Thorac. Soc.* **11**, 619–627 (2014).
25. J. H. Pedersen, P. Tønnesen, and H. Ashraf, "Smoking cessation and lung cancer screening," *Ann. Transl. Med.* **4**, 157 (2016).
26. P. Sönnerfors et al., "New interactive 3D visualization technique used in pulmonary rehabilitation programme in COPD, a randomized controlled study," *Eur. Respir. J.* **48**, PA692 (2016).

27. R. L. Murray et al., "Yorkshire Enhanced Stop Smoking (YESS) study: a protocol for a randomised controlled trial to evaluate the effect of adding a personalised smoking cessation intervention to a lung cancer screening programme," *BMJ Open* **10**, e037086 (2020).
28. "Vanderbilt Lung Screening Program," https://www.vumc.org/radiology/lung (2022).
29. R. Gao et al., "Technical report: quality assessment tool for machine learning with clinical CT," https://doi.org/10.48550/arXiv.2107.12842 (2021).
30. J. Wasserthal et al., "Total Segmentator: robust segmentation of 104 anatomical structures in CT images," https://doi.org/10.48550/arXiv.2208.05868 (2022).
31. A. A. A. Setio et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge," 2017, https://luna16.grand-challenge.org/ (accessed 9 Aug. 2022).
32. H. Tang, C. Zhang, and X. Xie, "Automatic pulmonary lobe segmentation using deep learning," in *Proc. – Int. Symp. Biomed. Imaging*, pp. 1225–1228 (2019).
33. J. Hofmanninger et al., "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur. Radiol. Exp.* **4**, 1–13 (2020).
34. S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *J. Comput. Phys.* **79**, 12–49 (1988).
35. R. Whitaker et al., "Segmentation of biological volume datasets using a level-set framework," in *Volume Graphics 2001. Eurographics*, K. Mueller and A. E. Kaufman, Eds., Springer, Vienna (2001).
36. C. Li et al., "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.* **19**, 3243–3254 (2010).
37. R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed., Springer International Publishing, New York (2022).
38. R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: a level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 158–175 (1995).
39. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
40. B. M. Smith et al., "Pulmonary emphysema subtypes on computed tomography in smokers," *Am. J. Med.* **127**, 94.e7–94.e23 (2014).

Thomas Z. Li is an MD/PhD candidate at Vanderbilt University in the Medical-Image Analysis and Statistical Interpretation Lab. He studied computer science at Duke University and has held software engineering roles in industry. His research is centered around combining AI with multi-modal medical data to advance the management of pulmonary nodules and lung cancer.

Biographies of the other authors are not available.