This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

**Article type:** Original Research

# Hierarchical deep learning for Pulmonary structures segmentation in surgical planning

Yuwen Chang[1], Yang Litao[1], Yu Zhen[1], Liu Jize[3], Wang Xin[3], Zheng Chao[3], Ji Qingru[2], Xiong Jianhao[2], Cheng Xuelian[1], Ge Zongyuan[1]

*[1] Monash University, Clayton, VIC 3800, Australia*

*[2] Airdoc Technology Co., Beijing, China*

*[3] Shukun Technology Co., Beijing, China*

Address correspondence to Yang Litao, Litao.Yang@monash.edu; Ge Zongyuan, Zongyuan.Ge@monash.edu

**Received:** November 27, 2025 / **Revised:** April 1, 2026 / **Accepted:** April 18, 2026

Just Accepted

**Abstract**

**Objective:** Precise segmentation of pulmonary segments from chest CT is essential for the diagnosis and treatment of lung diseases, especially for procedures such as segmentectomy. However, pulmonary segments are abstract anatomical units whose accurate delineation depends on the reliable segmentation of tubular structures including bronchi and arteries. This study aims to develop a robust automated framework to address this clinical challenge. **Methods:** Given chest CT images as input, we propose a two-stage automated segmentation methodology. In the first stage, a UMaxLSTM network with a union loss function is developed to perform efficient segmentation of pulmonary tubular structures (bronchi and arteries) as well as lung lobes. These structures provide anatomical boundaries critical for distinguishing pulmonary segments. In the second stage, we introduce the MultiFuseUNet network, which incorporates the tubular structure segmentation from stage one as guiding information to achieve precise automatic segmentation of

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
*Health Engineering* **2026, 2,** 9460014

pulmonary segments. **Results:** Segmentation performance of bronchi and arteries was evaluated on two public datasets, while pulmonary segment segmentation was validated on a private dataset with high-quality expert annotations. Experimental results demonstrate that the proposed two-stage methodology effectively segments pulmonary anatomical structures, achieving approximately 85% Dice score in pulmonary segment segmentation. **Conclusion:** The proposed approach provides thoracic surgeons and pulmonologists with an accurate and reliable tool for understanding pulmonary segment anatomy, showing significant clinical value. It is particularly suitable for preoperative planning of segmentectomy, helping clinicians avoid damage to adjacent healthy tissues.

**Keywords:** lung diseases; pulmonary segments; UMaxLSTM; MultiFuseUNet

## 1 Introduction

Pulmonary diseases are common illnesses that pose a significant threat to human health, particularly following the global impact of the COVID-19 pandemic, where high mortality and morbidity rates call for improved diagnostic and treatment techniques [1, 2]. In this study, we focus on pulmonary structure segmentation from chest CT images, which serve as the input modality for all stages of the proposed framework. Segmenting bronchial structures enables quantification of morphological changes, aiding diagnosis of diseases like COPD, ARDS, bronchial stenosis, and obliterative bronchiolitis [3]. Likewise, pulmonary artery morphology varies widely, and segmentation of individual arteries facilitates the identification of imaging biomarkers for both thrombotic and non-thrombotic pulmonary vascular diseases [4, 5]. Moreover, precise segmentation of these structures is essential for segmental pulmonary resection, as accurate analysis of pulmonary segments depends on reliable extraction of bronchial and arterial structures. This is because pulmonary segments are not directly observable in chest CT images, but are defined by bronchovascular structures rather than explicit intensity boundaries. CT images alone are insufficient to accurately delineate segmental regions due to the lack of clear visual cues and significant anatomical variability. Clinically, three principal surgical approaches, including segmentectomy, lobectomy, and pneumonectomy, are routinely employed, with automatic pulmonary segmentation serving as a crucial anatomical foundation for segmental resection. Compared to lung or lobe removal, segmentectomy reduces operative time, blood loss, and better preserves lung function [6]. Anatomically, following Boyden's nomenclature [7], each lung comprises approximately 18 segmental bronchi (8 on the left, 10 on the right) that form the basis

Just Accepted

of pulmonary segments, with natural bronchial variations [8] leading to interindividual differences in their number and configuration. As illustrated in Fig. 1 (a), pulmonary segments align with bronchial branching patterns and are delineated by segmental bronchi and their branches supplying distinct lung regions; their clinical relevance is demonstrated in Fig. 1 (b) through a segmentectomy example. Segment boundaries are defined by the intricate vascular anatomy, where segmental bronchi and arteries constitute each segment, while intersegmental veins demarcate the boundaries [9]. Successful segmentectomy depends on accurately identifying segment margins to minimize recurrence, posing a surgical challenge [10].

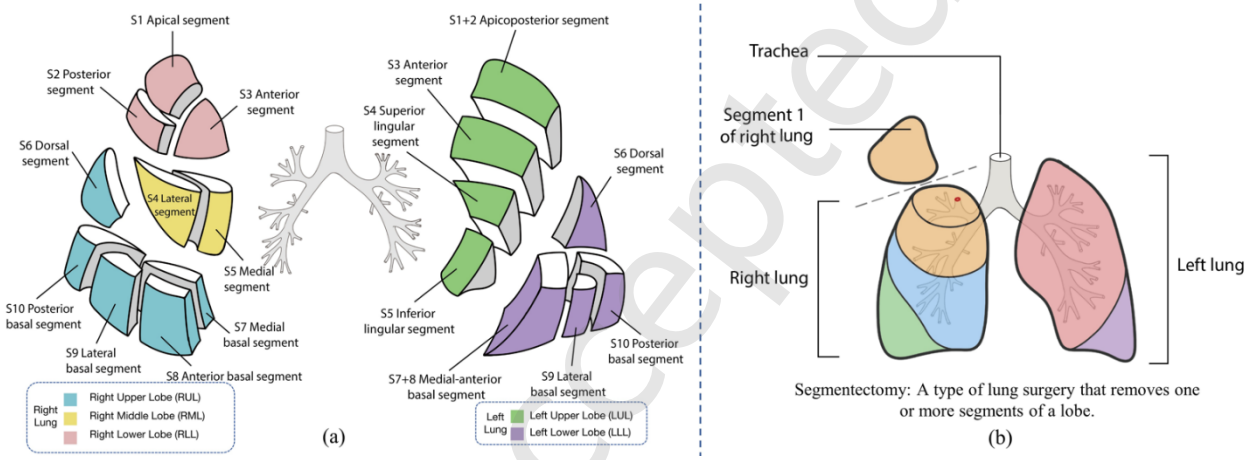


**Fig. 1.** Bronchopulmonary segmentation anatomy: (a) Detailed visualization of pulmonary segments (S1-S10) across right and left lungs, color-coded by lobar distribution (RUL, RML, RLL, LUL, LLL). (b) Simplified representation of lung anatomy showing tracheal bifurcation and lobar organization, with segmentectomy defined as surgical removal of one or more bronchopulmonary segments. We created the figure by Procreate and Illustrator.

Segmenting tubular structures such as bronchi and arteries poses substantial technical challenges due to their intricate branching, fine-scale details, and severe foreground–background imbalance, making them significantly more complex than conventional organ segmentation [11]. Bronchial, pulmonary artery, and lung segment segmentation are all critical tasks in thoracic image analysis. Existing methods often lack mechanisms to handle distal tubular structures or maintain boundary integrity, both essential for surgical success. Traditional methods for bronchial segmentation often struggle to capture distal branches due to weak contrast and complex topology [12]. More recent approaches have leveraged deep learning techniques, such as 3D CNNs [13] and graph neural networks [14], but they often struggle to capture long-range dependencies in complex distal regions. In addition, topology-aware losses such as clDice [15] and fuzzy topology

Just Accepted

constraints [16] aim to improve the segmentation of fine structures, yet they frequently suffer from instability and limited structural precision, as they rely on fragile skeleton extraction and provide sparse gradient signals, while lacking the ability to constrain fine-grained geometric details. Meanwhile, Pulmonary artery segmentation plays a key role in diagnosing vascular-related diseases such as pulmonary embolism and pulmonary hypertension [17, 18]. Recent works, including those from the PARSE Challenge, have adopted UNet-based architectures combined with structure-aware losses to address the challenges of vascular complexity [19,20]. Automated lung segment segmentation remains relatively underexplored compared to airway and vessel segmentation, and robust end-to-end, pixel-level methods are still lacking despite progress with implicit function-based approaches [21].

To address the critical challenges of structural ambiguity, segment boundary variability, and dependency on anatomical priors in both tubular structure and pulmonary segment segmentation, we present a comprehensive two-stage framework that explicitly leverages anatomical hierarchy to guide segmentation. Specifically, our method first employs a UMaxLSTM network, a purpose-built architecture that innovatively embeds Mamba and xLSTM modules within a UNet backbone, along with a novel loss to jointly segment bronchi and arteries, and then uses a MultiFuseUNet to perform pulmonary segment segmentation conditioned on the tubular structures obtained in the first stage. By combining a topology-aware union loss for robust delineation of intertwined bronchi and arteries with a structure-conditioned MultiFuseUNet for segment-level prediction, our framework effectively addresses both the segmentation difficulty of tubular structures and the anatomical variability inherent in pulmonary segments. The two-stage design is essential for this task due to the hierarchical anatomical dependencies: pulmonary segments are anatomically defined based on bronchial and arterial territories. By first delineating these tubular structures, our framework provides precise anatomical priors to guide the subsequent segment-level prediction. Our contributions can be briefly summarized as follows:

- We propose a two-stage framework that segments bronchi and arteries to anatomically guide pulmonary segment delineation, aligning with clinical practice where these structures define segment boundaries and support procedures like segmentectomy and preoperative planning.
- In the first stage, we develop UMaxLSTM, a novel architecture integrating Mamba and xLSTM modules to improve sequential modeling for tubular structures, and introduce Union Skeleton-aware Loss (USL) to enhance connectivity and distal robustness.
- In the second stage, we design a hybrid loss with anatomical constraints, combining region-based and boundary-aware terms to accommodate the fine-grained structural variability of pulmonary segments under complex anatomical conditions.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering*** **2026, 2,** 9460014

## 2 Related Work

### 2.1 Pulmonary Structures Segmentation

For bronchial segmentation tasks, previous studies have primarily relied on traditional bronchial segmentation methods, such as region growing, adaptive thresholding, and filter-based enhancement techniques [12, 22, 23]. However, as bronchi continue to branch, the intensity contrast between the bronchial walls and the airway lumen becomes increasingly weak, and the branches become progressively finer. Traditional methods often struggle to capture the thin, distal parts of the bronchial tree, limiting segmentation to the larger, proximal sections. In recent years, numerous studies have explored airway segmentation based on convolutional neural networks (CNNs) [13, 14, 20, 24-29]. Three-dimensional (3D) CNNs have been applied to bronchial segmentation using fixed-stride sliding windows and dynamically defined volumes of interest (VOIs) [28, 30]. To account for tubular topology awareness [25], spatially recursive convolutional layers and radial distance loss have been proposed. To better understand the intrinsic structure of bronchi, Qin et al. [13] proposed viewing bronchial segmentation as a 26-neighborhood connectivity prediction task. Graph neural networks have also been employed to integrate neighborhood knowledge, thereby enhancing feature extraction [14]. Qin et al. [11] introduced an attention distillation module and feature calibration to improve feature extraction for fine tubular structures. Additionally, specific loss functions have been proposed for bronchial tubular structures. For example, Shit et al. [15] introduced the centerline Dice (clDice) loss, which approximates the centerline of volumetric data using 3D skeletonization methods [31], although obtaining true values remains challenging. Nan et al. [16] designed a continuity and accumulation mapping (CAM) loss to enhance the continuity of bronchial structures. For pulmonary artery segmentation tasks, the pulmonary artery is a multi-level vascular system composed of the left and right main pulmonary arteries, the pulmonary trunk, and numerous segmental branches. Diseases such as pulmonary embolism [17] and pulmonary hypertension (PH) [32] are closely related to the pulmonary artery, which transports deoxygenated blood from the heart. Segmentation of the pulmonary artery structure is essential for the diagnosis and treatment of lung-related diseases. In the PARSE Challenge [19], various teams focused on artery segmentation primarily using the UNet architecture and its variants, incorporating additional information to better capture the features of tubular structures. Techniques include General Union Loss [20], weighted cross-entropy loss, and fine-tuning with clDice [15].

Although deep learning has achieved significant success in medical image segmentation [33], including the segmentation of lung structures such as lobes, airways, and vessels [11,34-36], prior

Just Accepted

studies have also explored related tasks such as lung lobe and lesion segmentation for clinical assessment [37] and pulmonary fissure segmentation using image filtering and machine learning techniques [38]. However, these methods are primarily designed for lobe-level or boundary-level analysis and do not directly address pulmonary segment segmentation. Research on lung segment segmentation, especially the automatic segmentation of variant lung segments, remains very limited. The potential of implicit functions in reconstructing lung segments has been explored [21], but to date there has not been a standard, universal, simple, and effective segmentation algorithm for variant lung segments when variations occur.

### 2.2 Mamba and xLSTM

Recent advances in sequential modeling architectures have introduced powerful frameworks that can effectively capture long-range dependencies in various domains, including medical imaging. State Space Models (SSMs) are one such approach, designed to model sequences through system dynamics represented by latent hidden states and parameter matrices. These models can perform complex computations at each time step and enable parallelized training. However, traditional SSMs face limitations such as high memory consumption and vanishing gradients. To overcome these issues, Structured State Space Sequence Models (S4) [39] restructure the dynamics of SSMs using the HIPPO framework [40], allowing improved long-range reasoning. S4 has demonstrated strong performance on benchmarks like the Long Range Arena [41], outperforming Transformer-based models [42].

Mamba [43] further improves upon S4 by introducing input-dependent mechanisms that dynamically select relevant input features. It also incorporates a hardware-efficient scanning algorithm that scales linearly with sequence length, leading to significantly faster computation during training and inference. Mamba has shown strong performance in modeling discrete data such as text and genomic sequences. Another recent development, xLSTM [44], extends traditional LSTM architectures and has demonstrated impressive results in Large Language Models (LLMs), outperforming both Transformers [42] and SSMs [39]. In computer vision, Vision Transformer (ViT) [45] and Vision Mamba [46] have been successful adaptations of these sequential models. Vision LSTM (ViL) [47] utilizes xLSTM blocks to enable sequence modeling in vision tasks. To handle the non-autoregressive nature of image data, ViL alternates the processing direction of mLSTM blocks [44] across layers. A notable advantage of ViL over ViT is its computational efficiency. Unlike the quadratic complexity of self-attention mechanisms in Transformers, ViL offers linear complexity in both computation and memory usage.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

# 3 Methods

## 3.1 Pulmonary tubular structure segmentation task

### 3.1.1 Preliminary

The problem setup involves segmenting tubular structures (bronchi and arteries) from CT scans, which is essential for accurate pulmonary analysis and diagnosis. Mathematically, let $X \in \mathbb{R}^{\{c_i \times N\}}$ represent the input CT scan volume, where $c_i$ is the number of input channels and $N = w \times h \times d$ denotes the total number of voxels. The ground truth tubular structure segmentation is represented by $G \in \{0,1\}^N$, where each voxel is labeled as either foreground (tubular structure) or background.

Our proposed UMaxLSTM network architecture, as illustrated in Fig. 2 (b), integrates sequential modeling capabilities with spatial feature extraction to effectively capture the complex topology of vascular and bronchial structures. The network processes the input volume $X$ to generate a probabilistic segmentation output $Y \in \mathbb{R}^{\{c_o \times N\}}$, where $c_o$ represents the number of output channels. The final binary segmentation mask $V$ is obtained by thresholding $Y$, assigning each voxel a value of 1 for the foreground structure (e.g., vessels or bronchi) and 0 for the background.

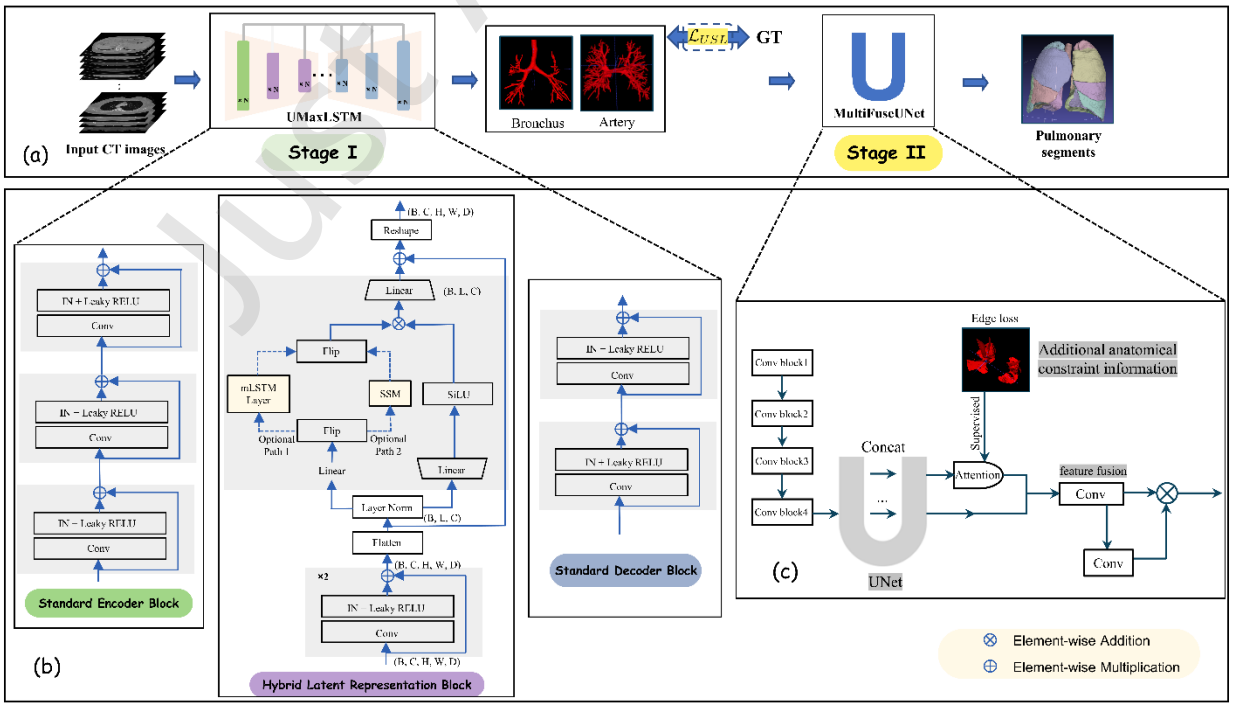

**Fig. 2.** Our comprehensive framework architecture. (a) Illustrates the overall workflow of our proposed network, where input images are processed through Stage I, consisting of the UMaxLSTM network and $L_{USL}$ to obtain vascular structure segmentation results, followed by Stage II, employing MultiFuseUNet to generate pulmonary segment segmentation outcomes. The detailed architectural frameworks of UMaxLSTM and MultiFuseUNet are presented in (b) and (c), respectively.

### 3.1.2 Task-Specific UMaxLSTM Configurations

For different anatomical structure segmentation tasks in medical images, we implemented task-specific optimizations to the UMaxLSTM network architecture, employing distinct sequence modeling layer configurations to accommodate the morphological and topological characteristics of various anatomical structures.

In arterial segmentation tasks, we adopted a "shallow convolution–middle LSTM–deep Mamba" configuration strategy: the first four stages $s < 4$ utilize basic residual convolutional blocks (BasicBlockD) for local feature extraction, the fifth stage $4 \leq s < 5$ incorporates Vision LSTM layers (ViLLayer) for medium-range spatial dependency modeling, and the deepest stages $s \geq 5$ deploy Mamba layers to capture global long-range dependencies. This design is motivated by the relatively consistent tubular morphology of arteries, which typically exhibit gradual tapering and fewer abrupt directional changes. Such structures are well-suited to extended local feature extraction before introducing sequential modeling. The use of LSTM and Mamba in later stages allows the network to progressively build contextual understanding while maintaining computational efficiency.

Conversely, for bronchial segmentation tasks characterized by intricate branching structures, we implemented an earlier introduction of sequence modeling capabilities: basic residual convolutional blocks are used only in the first three stages $s < 3$, Vision LSTM modeling is introduced from the fourth stage $3 \leq s < 4$, and Mamba models are employed in the fifth and deeper stages $s \geq 4$. This architectural design is fundamentally motivated by the complex hierarchical structure of bronchial trees, which exhibit significant variations in diameter, orientation, and connectivity across multiple scales. Such complexity necessitates earlier incorporation of sequence modeling (starting from stage 3) to better capture discontinuities and fine-scale spatial dependencies, given the extensive bifurcations and thin peripheral branches observed in bronchial anatomy. By tailoring the network architecture to the structural characteristics of each target anatomy, the UMaxLSTM framework achieves improved adaptability and segmentation accuracy across different segmentation tasks.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering*** **2026, 2,** 9460014

### 3.1.3 Union Skeleton-aware Loss (USL)

Following the established protocol in nnU-Net [48], all baseline experiments utilize the standard combination of dice loss and cross-entropy loss as the optimization objective:

$$L_{standard} = \alpha L_{dice} + \beta L_{CE} \quad (1)$$

where $\alpha = \beta = 1$ in the standard configuration.

For our proposed UMaxLSTM architecture, we introduce a novel Union Skeleton-aware Loss (USL) function that enhances the network's ability to capture long-range dependencies and structural continuity in complex anatomical structures. This specialized loss function leverages skeletal information to guide the segmentation process of intricate vascular and bronchial tree structures in 3D medical images, with a particular emphasis on preserving topological characteristics and penalizing structural inconsistencies in tubular formations.

The final loss function for UMaxLSTM incorporates both the standard loss and our proposed union loss with adaptive weighting:

$$L_{total} = L_{standard} + \omega L_{clDice} + \gamma L_{USL} \quad (2)$$

For all experiments, the value of $\omega$ was set to 1. For bronchial segmentation tasks, $\gamma$ was uniformly set to 0.1 across all network architectures. For arterial segmentation tasks, $\gamma$ was set to 1.0 for nnUNet, UxLSTM, and UMaxLSTM architectures, while a value of 0.2 was employed for UMambabot. These hyperparameter configurations were determined through preliminary experiments to optimize segmentation performance.

The Union Skeleton-aware Loss (USL) formulation is detailed in Algorithm 1. Following the pipeline illustrated in Algorithm 1, we formulate the final loss $L_{USL}$ based on a pair of structurally aware loss terms. Specifically, in step 6, we adopt the Generalized Union Loss (GUL) proposed by Zheng et al. [20] and adapt it to our skeletal representation. Here, $P$ is the predicted probability map, $T$ is the supervision target, and $W$ is a spatial weight map derived from either prediction or ground truth. The hyperparameters are chosen as follows: $\epsilon = 10^{-4}$ (stability term), $\delta = 1.0$ (smoothing constant), $\gamma = 0.7$ (nonlinear modulation), and $(\alpha, \beta) = (0.1, 0.9)$ (balancing factors for prediction and target).

As defined in step 7, the final Union Skeleton-aware Loss $L_{USL}$ combines two complementary GUL terms. The precision-weighted loss uses the predicted skeleton importance $\Psi_I$ as both

Just Accepted

supervision and weight, penalizing inconsistency with the ground truth distance map $\Phi_D$. Conversely, the sensitivity-weighted loss emphasizes ground truth skeleton regions via $\Phi_I$, supervising the predicted distance map $\Psi_D$ with higher weights in thinner and topologically sensitive regions. These two terms collectively guide the network to better capture fine-grained skeletal structures in both prediction and learning.

---

**Algorithm 1:** Computation of Union Skeleton-aware Loss (USL)

---

**Input:** Segmentation mask $M$, skeleton $S$, prediction $P$, predicted skeleton $S_P$

**Output:** Final loss $L_{USL}$

1: Compute Euclidean distance transform: $D = EDT(M)$

2: Construct skeleton radius map:

$$R_{S(x)} = \begin{cases} D(x), \ x \in S \\ 0, \ otherwise \end{cases}$$

3: Compute $R_{max} \leftarrow max(max(R_S), 1)$; $R_{min} \leftarrow max(min(R_S), 1)$

4: Compute ground truth weight maps:

$$\Phi_{D(x)} \leftarrow \frac{min(D(x), R_{max})}{R_{max}} \cdot \mathbb{I}(x \in M)$$

$$\Phi_{R(x)} \leftarrow \frac{R_{S(x)}}{R_{max}} \cdot \mathbb{I}(x \in M)$$

$$\Phi_{I(x)} \leftarrow \left(\frac{R_{max} - R_{S(x)} + R_{min}}{R_{max}}\right)^2 \cdot \mathbb{I}(x \in S)$$

5: Compute prediction weight maps $\Psi_D, \Psi_I, \Psi_R$ by replacing $\mathbb{I}(\cdot)$ with soft prediction $P(x)$

6: Define Generalized Union Loss:

$$L_{GU}(P, T, W) = 1 - \frac{\sum W \cdot (P + \epsilon)^{\gamma} \cdot T + \delta}{\sum W \cdot (\alpha \cdot P + \beta \cdot T) + \delta}$$

7: Compute final loss:

$$L_{USL} \leftarrow L_{GU}(\Psi_I, \Phi_D, \Psi_I) + L_{GU}(\Psi_D, \Phi_I, \Phi_I)$$

Just Accepted

8: Return $L_{USL}$

---

### 3.2 Pulmonary segment segmentation task

#### 3.2.1 Preliminary

The problem setup involves segmenting lung lobes from CT scans, a crucial step in diagnosing and treating pulmonary diseases. Mathematically, let $X$ represent the input CT scan image, which includes a series of anatomical constraints such as bronchi, arteries and edge information. $Y$ represents the corresponding ground truth lung lobe segmentation map. The objective is to find a mapping function $f : X \rightarrow Y$, where $f$ accurately segments the lung lobes from the input CT scan image $X$.

#### 3.2.2 The MultiFuseUNet

The overall architecture of our proposed MultiFuseUNet is illustrated in Fig. 2 (c). The input first passes through a multichannel head module consisting of four sequential convolutional blocks. The first two use $3 \times 3 \times 3$ grouped convolutions to increase channels (from $2 \rightarrow 4$, then $4 \rightarrow 16$), followed by a $1 \times 1 \times 1$ convolution reducing channels back to 4, and a final layer outputting a single-channel feature map. Each layer is followed by batch normalization and ReLU activation, enabling effective multi-scale feature extraction.

The processed features are then passed through a UNet with four downsampling and four upsampling layers, using $3 \times 3 \times 3$ convolutions throughout. Feature maps are progressively compressed and abstracted in the encoder, then upsampled and fused with skip connections in the decoder. The upsampled features are passed to a boundary attention module (3D convolution) that emphasizes structural boundaries.

Next, a feature fusion module combines outputs from the boundary attention and upsampling paths via concatenation. It includes two branches: one generates a weight map (through convolution → ReLU → Sigmoid) for highlighting important regions, while the other refines features through a separate convolution. The two outputs are fused by element-wise multiplication and addition to produce the final fused feature map. All convolutions in this module use $1 \times 1 \times 1$ kernels.

#### 3.2.3 Loss function

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

In the process of optimizing the lung CT segmentation network, we proposed a total loss function composed of four terms to enhance the segmentation performance of lung segments, expressed as follows:

$$L_{total} = L_{CE} + L_{Dice} + L_{IoU} + L_{EdgeII} \tag{3}$$

The total loss combines region-based terms ($L_{CE}$, $L_{Dice}$, $L_{IoU}$) with an edge-focused component $L_{EdgeII}$, which refines segmentation contours through a coarse-to-fine strategy. The weighting coefficients are set to $\omega = 0.6$ and $\sigma = 0.4$ respectively. The edge loss is defined as:

$$L_{EdgeII} = \omega L_{EdgeCoarse} + \sigma L_{EdgeRefine} \tag{4}$$

The coarse edge loss is designed to extract edge features:

$$L_{EdgeCoarse} = -\frac{1}{N}\sum_{i=1}^{N}\left(\left(\frac{1}{1 + e^{-\frac{E_i}{t}}}\right)\log(z_i) + \left(1 - \frac{1}{1 + e^{-\frac{E_i}{t}}}\right) * \log(1 - z_i)\right) \tag{5}$$

Where $N$ is the pixel count, $E_i$ is the edge ground truth, $z_i$ is the predicted probability, and $t$ is a temperature parameter controlling edge sensitivity.

The edge refinement loss focuses on challenging edges:

$$L_{EdgeRefine} = \sum_{i=1}^{N}\left(\eta \cdot \left(1 - \frac{\left(g_{vi[0]} \cdot p_{vi[0]} + g_{vi[1]} \cdot p_{vi[1]}\right)}{|g_{vi}| \cdot |p_{vi}|}\left(|g_{vi}| \cdot |p_{vi}|\right)\right)\right) \cdot E_i \tag{6}$$

Where $\eta = 0.5$ is a weight factor, $g_{vi}$ and $p_{vi}$ are ground truth and predicted edge vectors, and $E_i$ is the edge error. This uses cosine similarity to assess alignment between predicted and actual edges.

## 4 Experiments

### 4.1 Data Acquirement

#### 4.1.1 Airway Tree Modeling Challenge 2022 (ATM 22)

For bronchial segmentation, we used the ATM 22 dataset [54] consisting of 500 chest CT scans from Shanghai Chest Hospital and the public LIDC-IDRI dataset [55]. The CT scans were acquired using three devices: Philips iCT 256, GE Healthcare LightSpeed 16, and TOSHIBA

Just Accepted

Aquilion. All data were anonymized. Axial dimensions of scans are 512x512 pixels, slice thickness ranges from 0.450 to 1.000 mm, and slice counts vary between 157 and 1125. Annotations were performed by experienced radiologists. Although the dataset officially contains 500 cases, the ground truth annotations for the official test set are not publicly available. To enable offline evaluation and comparison of all models, we utilized the official training set (300 labeled cases), which was provided in two batches of 150 cases each. We randomly selected one batch (150 cases) as our experimental dataset, which was further randomly split into training and testing sets (80:20). Data were preprocessed in nnUNet format and used for network training.

#### 4.1.2 Pulmonary Artery Segmentation Challenge (Parse 22)

For the pulmonary artery segmentation task, we used the PARSE dataset [19], which consists of 203 chest CT images collected from multiple sites using Siemens and Philips devices. The sites include Harbin Medical University Cancer Hospital, Chinese PLA General Hospital, Hongqi Hospital Affiliated to Mudanjiang Medical University, and the First Affiliated Hospital of Jiamusi University. All data have been anonymized, with the axial dimensions of the CT images being 512x512 pixels and a thickness of 1 mm, with the number of slices ranging from 228 to 408. Annotations for the CT scans were performed by experienced radiologists. Since the ground truth labels for the official test set are not publicly available, we exclusively utilized the 100 labeled cases from the official training set to facilitate offline evaluation. These 100 samples were randomly divided into internal training and testing sets in an 80:20 ratio, preprocessing the data according to the nnUNet format and feeding it into the network for training.

#### 4.1.3 Lung segments Dataset

To evaluate the model's generalization capability on diverse clinical data, we utilized a large-scale in-house dataset comprising 718 chest CT scans (randomly split into 638 for training and 80 for testing). This dataset was retrospectively collected from multiple clinical centers via a qualified third-party data service. Prior to our access, all data were strictly de-identified to remove Protected Health Information (PHI) in accordance with ethical regulations, and the requirement for specific Institutional Review Board (IRB) approval was waived due to the retrospective and fully anonymized nature of the study. The dataset is characterized by high heterogeneity in acquisition protocols, reflecting real-world clinical variations; the scans were acquired using various scanner manufacturers (including GE, Siemens, and Philips) with slice thicknesses ranging from approximately 0.6 to 2.0 mm and in-plane resolutions varying between 0.5 and 0.9 mm. To ensure the reliability of the ground truth, the annotation process followed a rigorous two-stage pipeline: initial coarse segmentation was performed by trained annotators, followed by a pixel-level

Just Accepted

verification and refinement by a senior radiologist with over 10 years of experience in thoracic imaging. All scans were preprocessed by resampling to a unified isotropic resolution of 1.0 × 1.0 × 1.0 mm and intensity-normalized to a standard lung window range.

### 4.2 Evaluation Metrics

#### 4.2.1 Pulmonary tubular structure segmentation task

We employed seven evaluation metrics to comprehensively assess vascular structure segmentation performance, including Dice Similarity Coefficient (DSC), centerline Dice Similarity Coefficient (clDSC), Normalized Surface Distance (NSD), Precision, Recall, False Positive Rate (FPR), and Centerline Distance (CLD). Among them, DSC, clDSC, Precision, and Recall range from 0 to 1, where values closer to 1 indicate better overlap, topological correctness, and classification accuracy. NSD also ranges from 0 to 1, with higher values representing more accurate boundary alignment. In contrast, FPR and CLD are error-oriented metrics, where lower values indicate fewer false positives and better geometric alignment between predicted and ground truth centerlines, respectively.

#### 4.2.2 Pulmonary segment segmentation task

We evaluated segmentation accuracy using three specialized metrics based on the Dice Similarity Coefficient, ranging from 0 to 1 (higher values indicating better performance): DSC_s assesses the overall accuracy across lung segment classes; DSC_a evaluates arterial segmentation within each segment class; and DSC_b measures bronchial segmentation precision within each segment class. These complementary metrics are particularly relevant as lung segments are defined by their contained vascular structures rather than by distinct anatomical boundaries.

### 4.3 Implementation details

#### 4.3.1 Pulmonary tubular structure segmentation task

We adopted the default preprocessing of nnU-Net [48] and applied deep supervision in all experiments. Optimization was performed using SGD with momentum 0.99, weight decay 3e-5, and Nesterov acceleration. The initial learning rate was 0.001 for bronchial and 0.01 for arterial segmentation, later unified to 0.001 with union loss. A polynomial learning rate decay was applied. Mixed precision training and gradient clipping (threshold 12) were used for stability and efficiency. All models were trained for 50 epochs on a single 24GB NVIDIA A5000 GPU. Due to memory

Just Accepted

limits, UMambaEnc and SwinUNETR used a batch size of 1; others used a batch size of 2. Remaining settings followed default configurations. The total end-to-end training time was approximately 6 hours under these settings.

#### 4.3.2 Pulmonary segment segmentation task

The total training time is approximately three days, with inference latency on the order of seconds. Currently, annotating a complete case (comprising bronchi, arteries, and pulmonary segments) is extremely labor-intensive: outsourced manual annotation typically requires 1-2 days, while hospital-based semi-automatic workflows (AI + manual correction) still average around 2 hours. In contrast, our seconds-scale inference offers a dramatic efficiency leap. This speed is clinically sufficient, as the model serves as an auxiliary tool to assist physicians in structural analysis, with the final judgment remaining with the radiologist. Thus, such latency fits seamlessly into routine workflows.

## 5 Results

### 5.1 Pulmonary tubular structure segmentation task

Quantitative and qualitative comparisons of tubular structure segmentation are summarized in Table 1 and Figures 3-4. We evaluate eight representative segmentation networks on the bronchus and artery datasets under identical training conditions (50 epochs), including recent architectures based on Transformer, Mamba, and LSTM designs. Among all compared methods, the proposed UMaxLSTM demonstrates competitive and consistently strong performance across both bronchus and artery segmentation tasks, achieving leading results in multiple key metrics while maintaining stable performance across datasets.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

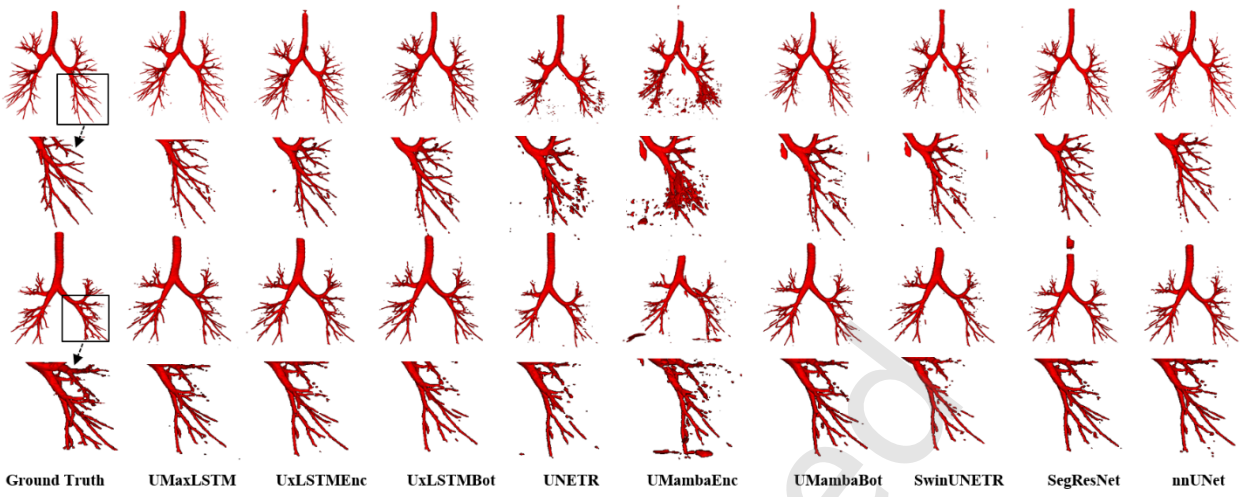


**Fig. 3.** Visual demonstration of segmentation results across different networks. For the comparative analysis of various networks, we selected two bronchial cases to demonstrate the segmentation outcomes. The lower right regions highlighted by black bounding boxes in the first and third rows are presented in detail in the second and fourth rows respectively for enhanced visualization. Notably, our proposed UMaxLSTM network achieves better segmentation performance despite being trained for merely 50 epochs.

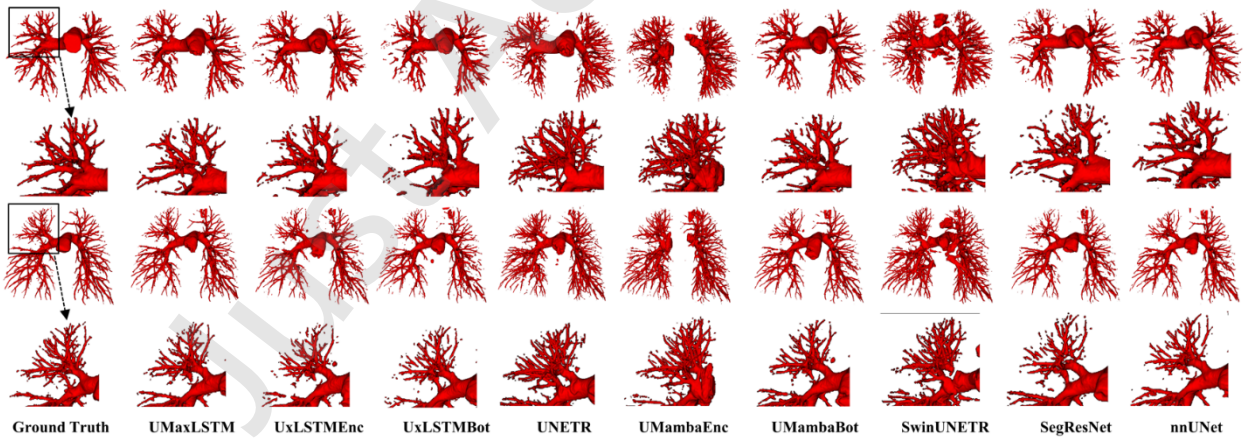


**Fig. 4.** Visual demonstration of segmentation results across different networks. For the comparative analysis of various networks, we selected two arterial cases to demonstrate the segmentation outcomes. The upper left regions highlighted by black bounding boxes in the first and third rows represented in detail in the second and fourth rows respectively for enhanced visualization. Notably, our proposed UMaxLSTM network achieves better segmentation performance despite being trained for merely 50 epochs.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

**Table 1.** Performance comparison of various networks on the Bronchus and Artery datasets under equal training conditions (50 epochs). All metrics are reported as percentages (%), except for the CLD metric, which is presented in its original numerical form.

| Model | Bronchus Dataset | | | | | | | Artery Dataset | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DSC (↑) | clDSC (↑) | NSD (↑) | Prec (↑) | Recall (↑) | FPR (↑) | CLD (↓) | DSC (↑) | clDSC (↑) | NSD (↑) | Prec (↑) | Recall (↑) | FPR (↑) | CLD (↓) |
| nnUNet [46] | 88.52 | 84.96 | 82.72 | 89.29 | 88.15 | 9.33e-05 | 1.31 | 82.81 | 75.56 | 61.53 | 85.44 | 81.45 | 5.55e-04 | 2.53 |
| SegResNet [47] | 88.88 | 87.13 | 83.56 | 87.90 | 90.08 | 1.03e-04 | 1.32 | 80.36 | 69.46 | 54.90 | 83.56 | 78.36 | 6.10e-04 | 2.96 |
| SwinUNETR [48] | 81.70 | 70.00 | 70.71 | 84.97 | 79.29 | 1.37e-04 | 2.02 | 57.12 | 58.05 | 40.47 | 51.82 | 61.80 | 2.29e-03 | **1.67** |
| UMambaBot [49] | 88.87 | 84.29 | 84.29 | 88.12 | 89.99 | 1.10e-04 | 1.20 | **83.87** | 83.15 | 64.11 | 85.98 | 82.97 | 5.40e-04 | 1.98 |
| UNETR [50] | 83.40 | 65.26 | 66.69 | 84.66 | 82.42 | 1.33e-04 | 3.93 | 71.55 | 57.81 | 43.14 | 68.91 | 75.39 | 1.37e-03 | 1.92 |
| UMambaEnc [49] | 72.69 | 58.09 | 54.34 | 77.27 | 69.55 | 1.83e-04 | 4.99 | 48.94 | 57.82 | 41.14 | 50.67 | 48.19 | 1.88e-03 | 1.79 |
| UxLSTMEnc [51] | 89.28 | 86.44 | **84.84** | 88.57 | 90.28 | 1.13e-04 | 1.22 | 82.46 | 73.65 | 59.11 | 86.72 | 79.59 | 4.75e-04 | 2.57 |
| UxLSTMBot [51] | 88.08 | 86.00 | 83.45 | 87.28 | 89.22 | 1.27e-04 | 1.27 | 82.26 | 73.60 | 58.75 | 86.52 | 79.40 | 4.85e-04 | 2.80 |
| UMaxLSTM | **89.55** | **87.67** | 84.53 | **90.96** | **90.95** | **7.67e-05** | **1.03** | 83.86 | **83.97** | **64.28** | **86.82** | **83.06** | **4.65e-04** | 1.75 |

Specifically, on the bronchus dataset, UMaxLSTM achieves the highest DSC (89.55%), clDice (87.67%), and Precision (90.96%), while maintaining the lowest false positive rate (7.67e-05). On the artery dataset, the model also demonstrates strong accuracy (DSC: 83.86%, clDice: 83.97%, NSD: 64.28%) and low topological error (CLD: 1.75), reflecting its robustness in delineating finer vascular structures. Visual comparisons in Figures 3 and 4 further support these findings: UMaxLSTM produces more accurate and structurally coherent segmentations, particularly in peripheral regions where other methods exhibit fragmentation or under-segmentation.

Furthermore, the effectiveness of the proposed Unified Structure-aware Loss (USL) is validated by incorporating it into four backbone networks: nnUNet, UMambaBot, UxLSTMEnc, and UMaxLSTM. As reported in Tables 2 and 3, USL consistently enhances segmentation performance across both datasets. On the bronchus dataset, USL improves UMaxLSTM's DSC to 90.66%, clDice to 86.91%, and NSD to 89.49%. On the artery dataset, it yields DSC of 85.72%, clDice of 69.48%, and NSD of 86.95%. These results confirm the superiority of USL over existing loss functions in both volumetric accuracy and topological preservation. Collectively, the proposed UMaxLSTM architecture combined with USL establishes a robust and anatomically faithful framework for the segmentation of complex tubular structures.

**Table 2.** Comparison of loss functions across four models on the bronchus dataset under identical training conditions (50 epochs). All evaluation metrics (DSC, NSD, and CLDSC) are reported as percentages (%).

Just Accepted

| Loss Function | nnUNet | | | UMambaBot | | | UxLSTMEnc | | | UMaxLSTM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DSC | NSD | clDSC | DSC | NSD | clDSC | DSC | NSD | clDSC | DSC | NSD | clDSC |
| Baseline | 88.52 | 82.72 | 84.96 | 88.87 | 84.29 | 86.15 | 72.69 | 54.34 | 86.44 | 89.55 | 84.53 | 86.67 |
| clDice [15] | 89.08 | 85.15 | **88.41** | 89.72 | 85.54 | 88.30 | 89.80 | 86.15 | 87.85 | 90.53 | 86.30 | 89.15 |
| cbDice [54] | 88.12 | 81.61 | 83.90 | 86.40 | 79.96 | 82.07 | 88.78 | 83.08 | 85.66 | 88.32 | 83.02 | 85.84 |
| BDoU [55] | 89.34 | 83.29 | 83.82 | 86.45 | 81.27 | 79.10 | 89.40 | 84.45 | 85.88 | 89.97 | 84.17 | 85.20 |
| Dice_BDoU | 87.97 | 82.69 | 84.69 | 90.19 | 85.29 | 86.39 | 90.16 | 85.47 | 86.99 | 90.25 | 85.14 | 85.71 |
| CLMDC [54] | 88.37 | 81.21 | 83.51 | 85.54 | 80.60 | 81.23 | 89.11 | 83.13 | 84.73 | 87.78 | 82.45 | 85.17 |
| clDice_clCE [56] | 88.43 | 84.12 | 87.00 | 89.89 | 85.46 | 88.57 | 90.78 | 86.56 | 88.94 | 90.46 | 86.30 | 88.17 |
| USL (Ours) | **89.38** | **85.22** | 87.11 | **90.23** | **85.54** | **88.93** | **91.16** | **86.77** | **89.12** | **90.66** | **86.91** | **89.49** |

**Table 3.** Comparison of loss functions across four models on the artery dataset under identical training conditions (50 epochs). All evaluation metrics (DSC, NSD, and CLDSC) are reported as percentages (%).

| Loss Function | nnUNet | | | UMambaBot | | | UxLSTMEnc | | | UMaxLSTM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | DSC | NSD | clDSC | DSC | NSD | clDSC | DSC | NSD | clDSC | DSC | NSD | clDSC |
| Baseline | 82.81 | 61.53 | 75.56 | 83.87 | 64.11 | 79.87 | 82.46 | 59.11 | 73.65 | 83.86 | 64.28 | 83.97 |
| clDice [15] | 84.80 | 67.73 | 80.93 | 85.89 | **70.05** | 83.27 | 84.63 | 67.48 | 81.13 | 85.03 | 67.71 | 84.11 |
| cbDice [54] | 84.75 | 65.14 | 80.39 | 84.60 | 64.73 | 80.75 | 84.40 | 63.63 | 79.42 | 84.19 | 63.62 | 82.57 |
| BDoU [55] | 82.21 | 59.54 | 74.78 | 83.00 | 61.98 | 77.49 | 82.88 | 60.42 | 75.16 | 82.67 | 59.59 | 85.11 |
| Dice_BDoU | 81.94 | 58.08 | 72.77 | 83.73 | 62.36 | 76.39 | 83.54 | 61.49 | 76.53 | 83.80 | 63.12 | 84.57 |
| CLMDC [54] | 83.60 | 61.09 | 76.61 | 84.86 | 64.65 | 80.16 | 84.10 | 62.87 | 78.07 | 84.16 | 61.95 | 83.58 |
| clDice_clCE [56] | 84.88 | 66.93 | 80.82 | 85.65 | 69.28 | **84.53** | **85.05** | 67.97 | 81.82 | 85.31 | 68.10 | 84.11 |
| USL (Ours) | **85.00** | **68.45** | **82.12** | **85.71** | 69.69 | 84.00 | 85.00 | **68.12** | **82.01** | **85.72** | **69.48** | **86.95** |

## 5.2 Pulmonary segment segmentation task

Quantitative and qualitative results of pulmonary segment segmentation are presented in Figure 5, Figure 6, and Table 4. We conducted a systematic ablation study to evaluate the contribution of each anatomical constraint and supervision component by progressively removing the bronchial input, arterial input, and edge-guided loss (EdgeII) from the full MultiFuseNet architecture. As shown in Table 4, the complete MultiFuseNet model achieves the best overall performance across all evaluated metrics: 83.47% in segment-level Dice (DSC_s), 85.03% in artery-aligned Dice (DSC_a), and 86.13% in bronchus-aligned Dice (DSC_b), significantly outperforming the baseline UNet model (e.g., a gain of 3.98% in DSC_s over the bronchus-ablated

Just Accepted

variant). Notably, the exclusion of edge-aware supervision resulted in the largest performance degradation (DSC_s dropped from 83.47% to 74.52%), indicating that the incorporation of intersegmental venous information has the greatest impact on pulmonary segment delineation and underscoring the critical role of anatomical boundary refinement in enhancing spatial accuracy.

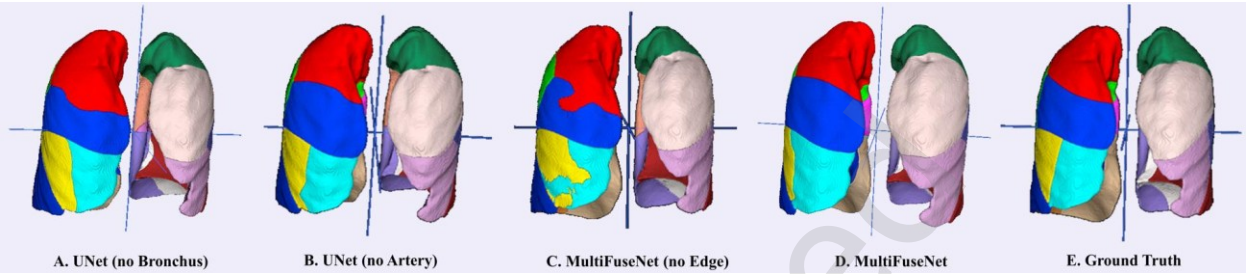


**Fig. 5.** Comparative visualization of pulmonary segment segmentation results across different deep learning architectures. (A) UNet without bronchial guidance; (B) UNet without arterial guidance; (C) MultiFuseNet without edge supervision; (D) Our proposed MultiFuseNet with full anatomical and edge constraints; (E) Ground truth. Our method achieves superior boundary delineation and more accurate preservation of distal pulmonary structures, demonstrating clear improvements over prior models.

**Table 4.** Performance Evaluation of MultiFuseNet. Superior Segmentation Accuracy via Feature Fusion and Loss Optimization Measured by Multiple DSC Metrics (%).

| Model | DSC_s (↑) | DSC_a (↑) | DSC_b (↑) |
|---|---|---|---|
| UNet [33] (no Bronchus) | 79.52 | 81.31 | 82.15 |
| UNet [33] (no Artery) | 81.22 | 82.71 | 83.96 |
| MultiFuseNet (no Edge) | 74.52 | 75.63 | 77.35 |
| MultiFuseNet | **83.47** | **85.03** | **86.13** |

Figure 5 illustrates segmentation outputs under different model configurations. Compared to the ablated variants, MultiFuseNet demonstrates more accurate preservation of anatomical boundaries, particularly in challenging regions such as the hilar zone and distal pulmonary segments, which are areas prone to over-segmentation or structural discontinuity. These visual improvements are most evident in the maintenance of fine segment margins and the reduction of cross-lobe leakage.

To contextualize these findings, Figure 6 depicts the anatomical correspondence between bronchial and arterial trees and their associated pulmonary segments. This structural interpretation

Just Accepted

substantiates the motivation for dual anatomical constraints and highlights the importance of aligning segmentation outputs with known anatomical priors. Collectively, our results confirm that integrating multi-source anatomical information with edge-aware supervision facilitates more accurate, consistent, and anatomically coherent pulmonary segmentations.

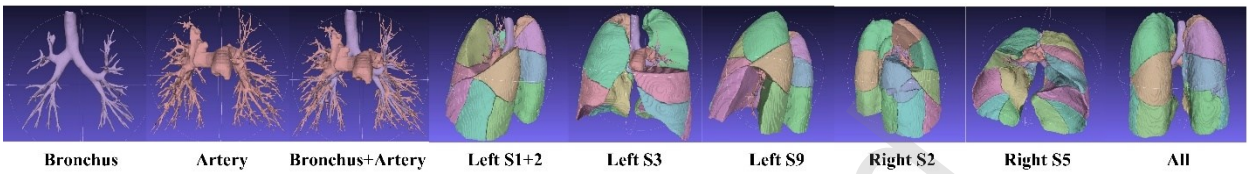


**Fig. 6.** To better illustrate their anatomical relationships, we reconstructed the bronchial arteries and pulmonary segments. Each pulmonary segment corresponds to a structural region defined by the partial vascular territories of both bronchial and arterial branches.

### 5.3 Further Analysis

Although both Mamba and xLSTM capture long-range dependencies, they function complementarily in our design: Mamba efficiently models global topology via state-space equations at deep stages, while xLSTM leverages matrix memory to enhance fine-grained feature retrieval at intermediate layers. Our ablation study (Table 5) empirically validates this choice, showing that the hybrid configuration consistently outperforms homogeneous stacking. Specifically, the optimal arrangement (using xLSTM at Stage 3 for Bronchus) achieves a DSC of 89.55%, demonstrating that strategically combining xLSTM's retrieval capability with Mamba's global abstraction is essential for accurate tubular segmentation.

To demonstrate robustness, we conducted a sensitivity analysis on the focusing parameter $\gamma$ and weighting factors $(\alpha, \beta)$, as detailed in Table 6. The results indicate that the model remains highly stable with negligible performance fluctuations ($< 0.2$) across a reasonable range $\gamma \in [0.5, 0.9]$. However, removing the non-linear focusing ($\gamma = 1.0$) or the specific imbalance weighting leads to noticeable degradation, empirically justifying the necessity of our default configuration.

**Table 5.** Quantitative comparison of different hybrid block configurations. This table evaluates the performance of various combinations of Convolutional, Mamba, and ViL layers applied to the Bronchus and Artery datasets. $S_i$ denotes the $i-th$ stage of the encoder.

| Configuration Strategy (Layer Assignment) | Bronchus Dataset | Artery Dataset |
|---|---|---|

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

| | DSC ( ↑ ) | clDSC ( ↑ ) | NSD ( ↑ ) | DSC ( ↑ ) | clDSC ( ↑ ) | NSD ( ↑ ) |
|---|---|---|---|---|---|---|
| $S_{0-2}$: Conv,$S_3$:Mamba,$S_{\geq 4}$: ViL | 89.73 | 87.11 | 84.08 | 82.67 | 75.23 | 61.45 |
| $S_{0-3}$: Conv, $S_4$: Mamba, $S_{\geq 5}$: ViL | 88.75 | 84.13 | 83.59 | 81.97 | 72.35 | 58.94 |
| $S_{0-3}$: Conv, $S_4$: ViL, $S_{\geq 5}$: Mamba | 87.89 | 87.13 | 83.09 | **83.86** | **83.97** | **64.28** |
| Mixed ($S_0$, $S_3$): Conv; $S_2$: Mamba; $S_{\geq 4}$: ViL; $S_{\geq 5}$: Mamba | 87.88 | 87.02 | 83.12 | 82.33 | 75.17 | 59.86 |
| $S_{0-2}$: Conv, $S_3$: ViL, $S_{\geq 4}$: Mamba | **89.55** | **87.67** | **84.53** | 83.46 | 83.01 | 62.49 |

As visualized in Figure 7, we provide a balanced evaluation of the framework's performance under varying conditions. The first two rows of Figure 7 illustrate the model's resilience to imperfect Stage I inputs. In clinical practice, tubular segmentation (bronchi/arteries) often suffers from topological breaks or misses distal branches. However, our MultiFuseUNet effectively leverages the global contextual information captured by the hybrid Mamba-xLSTM encoder to "bridge" these gaps. The results show that even with disconnected vascular priors (Col 1), the model generates smooth and accurate inter-lobar boundaries, confirming its robustness to local noise. Conversely, the bottom two rows demonstrate "worst-case" scenarios involving severe pathology, such as large tumors or consolidation. In these instances, the pathological changes completely obscure the vascular/bronchial signal, causing Stage I to generate highly fragmented or empty masks. Lacking sufficient anatomical guidance, the model struggles to define precise boundaries in the affected regions. This analysis suggests that while the model is robust to minor structural errors, its performance in extreme pathological cases remains dependent on the visibility of underlying anatomical landmarks.

**Table 6.** Hyperparameter sensitivity analysis of the proposed USL on the Artery dataset using the nnUNet baseline. We evaluate the model's robustness by varying the focusing parameter $\gamma$ and the balancing factors $(\alpha, \beta)$ around the default settings.

| Experiment Group | Parameter Setting | DSC ( ↑ ) | NSD ( ↑ ) | clDSC ( ↑ ) |
|---|---|---|---|---|
| Ours | $\gamma = 0.7, (\alpha, \beta) = (0.1, 0.9)$ | 85.00 | 68.45 | 82.12 |
| Focusing Param: Fixed $(\alpha, \beta) = (0.1, 0.9)$ | $\gamma = 0.5$ | 84.92 | 68.35 | 82.05 |
| | $\gamma = 0.9$ | 84.85 | 68.20 | 81.90 |
| | $\gamma = 1.0$ | 84.10 | 67.10 | 81.00 |
| Balancing Factor: Fixed $\gamma = 0.7$ | $(\alpha, \beta) = (0.3, 0.7)$ | 84.50 | 67.80 | 81.50 |
| | $(\alpha, \beta) = (0.5, 0.5)$ | 83.20 | 66.00 | 79.50 |

Just Accepted

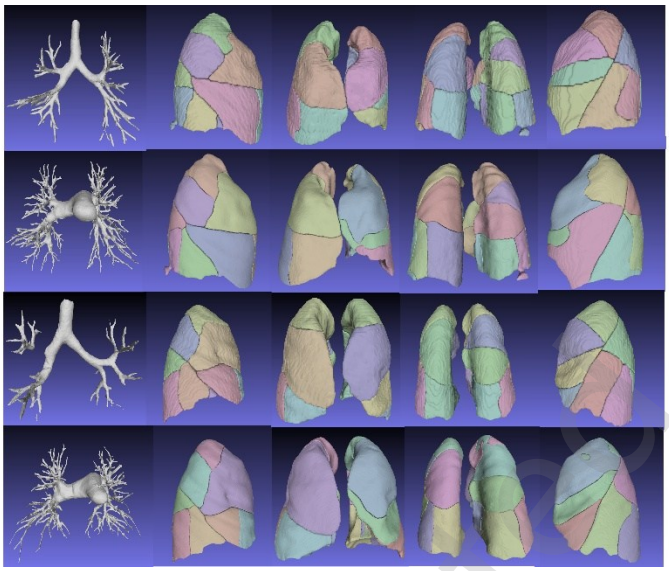

**Figure 7.** Qualitative visualization of model robustness and limitations. Rows 1-2 (Robustness): Despite topological breaks or missing branches in the Stage I priors (Col 1), the model successfully recovers complete lobe boundaries (Cols 2-5), demonstrating resilience to imperfect vascular guidance. Rows 3-4 (Failure Cases): Under severe pathology (e.g., tumors), obscured vascular structures lead to noisy Stage I inputs and subsequent boundary misalignments, highlighting the dependency on visible anatomical cues. Note: The specific colors assigned to distinct lung lobes are randomized by MeshLab for visualization purposes; the primary focus should be on the geometric shape and boundaries rather than the color itself.

To ensure a fair comparison, all models are evaluated in terms of parameters, FLOPs, and inference time under a unified input setting. We adopt the standardized patch sizes defined by nnU-Net after preprocessing, i.e., [128, 160, 112] for Bronchus and [96, 160, 160] for Artery. For models with specific input constraints, adjusted patch sizes are used only for complexity profiling and are indicated in Table 7. In particular, for SwinUNETR on Bronchus, the automatically planned patch size [128, 160, 112] does not satisfy the model's divisibility constraint; therefore, the nearest valid input size [128, 160, 128] is used for complexity profiling, and the corresponding results are reported in parentheses in Table 7. As shown in Table 7, different models exhibit varying trade-offs between computational cost and efficiency. While some methods achieve lower FLOPs or faster inference, they do not consistently balance both aspects. In contrast, the nnU-Net-based framework demonstrates a favorable balance, maintaining moderate inference time (e.g., 68.026 ms for Bronchus and 75.578 ms for Artery) despite relatively higher FLOPs. Importantly, the proposed method achieves better segmentation performance by incorporating anatomical priors and

Just Accepted

boundary supervision, which are essential for accurate pulmonary segment delineation. Given that inference times on the order of tens of milliseconds are acceptable in most clinical scenarios, the additional computational cost is justified. Overall, the proposed method provides a practical trade-off between accuracy and efficiency.

Table 7. Computational complexity comparison of different models on Bronchus and Artery datasets under standardized input settings.

| Model | Params (↓) | Bronchus Dataset | | Artery Dataset | |
|---|---|---|---|---|---|
| | | FLOPs(G) (↓) | Time(ms) (↓) | FLOPs(G) (↓) | Time(ms) (↓) |
| nnUNet [46] | 30.786 | 3965.388 | 68.026 | 4248.630 | 75.578 |
| SegResNet [47] | 18.796 | 633.371 | 111.559 | 678.612 | 122.239 |
| SwinUNETR [48] | 62.187 | N/A (979.134) | N/A (258.532) | 914.524 | 238.664 |
| UMambaBot [49] | 42.120 | 7320.923 | 152.816 | 7843.846 | 169.945 |
| UNETR [50] | 93.048 | 213.981 | 60.052 | 229.265 | 65.432 |
| UMambaEnc [49] | 42.856 | 7388.568 | 237.099 | 7916.327 | 258.571 |
| UxLSTMEnc [51] | 43.796 | 10893.502 | 198.885 | 11671.616 | 218.385 |
| UxLSTMBot [51] | 42.087 | 7320.978 | 154.483 | 7843.905 | 169.525 |
| UMaxLSTM | 46.929 | 11139.923 | 185.561 | 11935.636 | 202.456 |

## 6 Conclusion

In this work, we propose a two-stage framework for pulmonary segment segmentation, addressing the limitations of manual diagnosis and anatomical ambiguity. Our UMaxLSTM network, guided by a skeleton-aware loss, accurately segments complex bronchial and arterial structures, which are then used by MultiFuseUNet as anatomical priors for segment-level delineation. Experiments conducted on both public and private datasets show promising performance in segmenting bronchial and arterial structures, suggesting the potential applicability of our method in anatomically detailed pulmonary segment delineation. This approach holds strong potential for clinical applications such as preoperative planning, by enabling reliable, objective, and anatomically consistent pulmonary segment identification. Future work will explore unified training across both stages to further assess end-to-end robustness and clinical applicability.

However, we acknowledge certain limitations in the current framework. First, the model's reliance on topological priors makes it vulnerable in scenarios with severe pathology (e.g., large tumors or consolidation) where the underlying vascular structures are obliterated, leading to

Just Accepted

potential boundary errors. Second, while validated on heterogeneous data, the model's robustness on extremely low-resolution or low-dose CT scans remains to be fully quantified. Finally, the hybrid Mamba-xLSTM architecture, while effective for global modeling, introduces a higher computational overhead compared to lightweight CNNs.

Future work will explore unified training across both stages to further assess end-to-end robustness and clinical applicability, specifically aiming to optimize computational efficiency and develop pathology-aware adaptation strategies to address these challenges.

**Acknowledgements**

The authors gratefully acknowledge Shukun Technology for providing the datasets used in this study and for offering partial technical support.

**Author contribution statement**

Chang Yuwen conducted the entire workflow, including Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, and Visualization, and also contributed to Writing – original draft preparation. Jize Liu participated in the lung-segment–related part of the work. Litao Yang, Zhen Yu, Xin Wang, Chao Zheng, Qingru Ji, Jianhao Xiong, Xuelian Cheng, and Zongyuan Ge contributed to Resources, Supervision, and Writing – review & editing. All authors have read and approved the final manuscript..

**Consent**

Not applicable. All datasets used in this study were fully anonymized prior to analysis, and no additional informed consent was required.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering* 2026, 2,** 9460014

## Data availability

The publicly available datasets used in this study can be downloaded from their corresponding official websites. The private datasets collected by Shukun Technology are not publicly available due to institutional data-sharing restrictions but can be accessed from the corresponding author upon reasonable request.

## Declaration of competing interest

Jize Liu, Xin Wang, and Chao Zheng are employees of Shukun Technology, Beijing, China. Qingru Ji and Jianhao Xiong are employees of Airdoc, Beijing, China. The other contributing authors report no competing interests related to this work.

## Ethical approval

This study involved publicly available datasets as well as private datasets collected by Shukun Technology. The publicly available datasets were fully anonymized and obtained under the ethical approvals described in their original publications; therefore, no additional approval was required for their use in this study. The private datasets were de-identified, and their use in this research was approved by the institutional review board of Shukun Technology. No additional individual consent was required because all data were anonymized prior to analysis.

## Funding

None.

Just Accepted

This is a Just Accepted version, keep updated with https://doi.org/10.26599/HE.2026.9460014
***Health Engineering*** **2026, 2,** 9460014

## Use of AI statement

During the preparation of this work, the authors used ChatGPT for language polishing. After using this tool, the authors carefully reviewed and edited the content as needed and take full responsibility for the final version of the manuscript.

## References

[1] Adams SJ, Stone E, Baldwin DR, Vliegenthart R, Lee P, Fintelmann FJ. Lung cancer screening. *The Lancet*. 2023;401(10374):390–408.

[2] Gandhi Z, Gurram P, Amgai B, Lekkala SP, Lokhandwala A, Manne S, Mohammed A, Koshiya H, Dewaswala N, Desai R, et al. Artificial intelligence and lung cancer: impact on improving patient outcomes. *Cancers*. 2023;15(21):5236.

[3] Howling S, Evans T, Hansell D. The significance of bronchial dilatation on CT in patients with adult respiratory distress syndrome. *Clinical Radiology*. 1998;53(2):105–109.

[4] Zhou C, Chan H-P, Sahiner B, Hadjiiski LM, Chughtai A, Patel S, Wei J, Ge J, Cascade PN, Kazerooni EA. Automatic multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications. *Medical Physics*. 2007;34(12):4567–4577.

[5] Rahaghi FN, Ross JC, Agarwal M, González G, Come C, Diaz A, Vegas-Sánchez-Ferrero G, Hunsaker A, Estépar RSJ, Waxman A, et al. Pulmonary vascular morphology as an imaging biomarker in chronic thromboembolic pulmonary hypertension. *Pulmonary Circulation*. 2016;6(1):70–81.

[6] Handa Y, Tsutani Y, Mimae T, Miyata Y, Okada M. Postoperative pulmonary function after complex segmentectomy. *Annals of Surgical Oncology*. 2021;28:8347–8355.

[7] EA B. The intrahilar and related segmental anatomy of the lung. *Surgery*. 1945;18:706–731.

Just Accepted

[8] Wooten C, Patel S, Cassidy L, Watanabe K, Matusz P, Tubbs RS, Loukas M. Variations of the tracheobronchial tree: anatomical and clinical significance. *Clinical Anatomy*. 2014;27(8):1223–1233.

[9] Oizumi H, Kato H, Endoh M, Inoue T, Watarai H, Sadahiro M. Techniques to define segmental anatomy during segmentectomy. *Annals of Cardiothoracic Surgery*. 2014;3(2):170.

[10] Oizumi H, Kanauchi N, Kato H, Endoh M, Suzuki J, Fukaya K, Sadahiro M. Anatomic thoracoscopic pulmonary segmentectomy under 3-dimensional multidetector computed tomography simulation: a report of 52 consecutive cases. *The Journal of Thoracic and Cardiovascular Surgery*. 2011;141(3):678–682.

[11] Qin Y, Zheng H, Gu Y, Huang X, Yang J, Wang L, Yao F, Zhu Y-M, Yang G-Z. Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT. *IEEE Transactions on Medical Imaging*. 2021;40(6):1603–1617.

[12] Mori K, Hasegawa J-i, Suenaga Y, Toriwaki J-i. Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. *IEEE Transactions on Medical Imaging*. 2000;19(2):103–114.

[13] Qin Y, Chen M, Zheng H, Gu Y, Shen M, Yang J, Huang X, Zhu Y-M, Yang G-Z. AirwayNet: a voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. In: *MICCAI*. Springer; 2019:212–220.

[14] Selvan R, Kipf T, Welling M, García-Ulloa A, Pedersen JH, Petersen J, de Bruijne M. Graph refinement based airway extraction using mean-field networks and graph neural networks. *Medical Image Analysis*. 2020;64:101751.

[15] Shit S, Paetzold JC, Sekuboyina A, Ezhov I, Unger A, Zhylka A, Pluim JP, Bauer U, Menze BH. clDice: A novel topology-preserving loss function for tubular structure segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2021:16560–16569.

[16] Nan Y, Del Ser J, Tang Z, Tang P, Xing X, Fang Y, Herrera F, Pedrycz W, Walsh S, Yang G. Fuzzy attention neural network to tackle discontinuity in airway segmentation. *IEEE Transactions on Neural Networks and Learning Systems*. 2023.

Just Accepted

[17] Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, Huisman MV, Humbert M, Jennings CS, Jiménez D, et al. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism. *European Heart Journal*. 2020;41(4):543–603.

[18] Galiè N, Humbert M, Vachiery J-L, Gibbs S, Lang I, Torbicki A, Simonneau G, Peacock A, Vonk Noordegraaf A, Beghetti M, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *European Heart Journal*. 2016;37(1):67–119.

[19] Luo G, Wang K, Liu J, Li S, Liang X, Li X, Gan S, Wang W, Dong S, Wang W, et al. Efficient automatic segmentation for multi-level pulmonary arteries: The PARSE challenge. *arXiv preprint* arXiv:2304.03708; 2023.

[20] Zheng H, Qin Y, Gu Y, Xie F, Yang J, Sun J, Yang G-Z. Alleviating class-wise gradient imbalance for pulmonary airway segmentation. *IEEE Transactions on Medical Imaging*. 2021;40(9):2452–2462.

[21] Kuang K, Zhang L, Li J, Li H, Chen J, Du B, Yang J. What makes for automatic reconstruction of pulmonary segments? In: *MICCAI*. Springer; 2022:495–505.

[22] Van Rikxoort EM, Baggerman W, Van Ginneken B. Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach. In: *Proceedings of the Second International Workshop on Pulmonary Image Analysis*; 2009:341–349.

[23] Lo P, Van Ginneken B, Reinhardt JM, Yavarna T, De Jong PA, Irving B, Fetita C, Ortner M, Pinho R, Sijbers J, et al. Extraction of airways from CT (EXACT'09). *IEEE Transactions on Medical Imaging*. 2012;31(11):2093–2107.

[24] Garcia-Uceda Juarez A, Selvan R, Saghir Z, de Bruijne M. A joint 3D U-Net-graph neural network-based method for airway segmentation from chest CTs. In: *Machine Learning in Medical Imaging (MLMI 2019), MICCAI Workshop*; 2019:583–591.

[25] Wang C, Hayashi Y, Oda M, Itoh H, Kitasaka T, Frangi AF, Mori K. Tubular structure segmentation using spatial fully-connected network with radial distance loss for 3D medical images. In: *MICCAI 2019, Part VI*; 2019:348–356.

[26] Yun J, Park J, Yu D, Yi J, Lee M, Park HJ, Lee J-G, Seo JB, Kim N. Improvement of fully automated airway segmentation on volumetric CT images using a 2.5-dimensional convolutional neural network. *Medical Image Analysis*. 2019;51:13–20.

Just Accepted

[27] Charbonnier J-P, Van Rikxoort EM, Setio AAA, Schaefer-Prokop CM, Van Ginneken B, Ciompi F. Improving airway segmentation in CT using leak detection with convolutional networks. *Medical Image Analysis*. 2017;36:52–60.

[28] Garcia-Uceda Juarez A, Tiddens HA, de Bruijne M. Automatic airway segmentation in chest CT using convolutional neural networks. In: *RAMBO/BIA/TIA Workshops, MICCAI 2018*; 2018:238–250.

[29] Zhao T, Yin Z, Wang J, Gao D, Chen Y, Mao Y. Bronchus segmentation and classification by neural networks and linear programming. In: *MICCAI 2019, Part VI*. Springer; 2019:230–239.

[30] Meng Q, Roth HR, Kitasaka T, Oda M, Ueno J, Mori K. Tracking and segmentation of the airways in chest CT using a fully convolutional network. In: *MICCAI 2017, Part II*. Springer; 2017:198–207.

[31] Lee T-C, Kashyap RL, Chu C-N. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing*. 1994;56(6):462–478.

[32] Shahin Y, Alabed S, Alkhanfar D, Tschirren J, Rothman AM, Condliffe R, Wild JM, Kiely DG, Swift AJ. Quantitative CT evaluation of small pulmonary vessels has functional and prognostic value in pulmonary hypertension. *Radiology*. 2022;305(2):431–440.

[33] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *MICCAI 2015, Part III*. Springer; 2015:234–241.

[34] Gerard SE, Reinhardt JM. Pulmonary lobe segmentation using a sequence of convolutional neural networks for marginal learning. In: *ISBI 2019*. IEEE; 2019:1207–1211.

[35] Gerard SE, Patton TJ, Christensen GE, Bayouth JE, Reinhardt JM. FissureNet: A deep learning approach for pulmonary fissure detection in CT images. *IEEE Transactions on Medical Imaging*. 2018;38(1):156–166.

[36] Nardelli P, Jimenez-Carretero D, Bermejo-Pelaez D, Washko GR, Rahaghi FN, Ledesma-Carbayo MJ, Estépar RSJ. Pulmonary artery–vein classification in CT images using deep learning. *IEEE Transactions on Medical Imaging*. 2018;37(11):2428–2440.

[37] Khomduean P, Phuaudomcharoen P, Boonchu T, Taetragool U, Chamchoy K, Wimolsiri N, Jarrusrojwuttikul T, Chuajak A, Techavipoo U, Tweeatsani N. Segmentation of lung lobes and

Just Accepted

lesions in chest CT for the classification of COVID-19 severity. Scientific Reports. 2023;13(1):20899.

[38] Fufin M, Makarov V, Alfimov VI, Ananev VV, Ananeva A. Pulmonary fissure segmentation in CT images using image filtering and machine learning. Tomography. 2024;10(10):1645–1664.

[39] Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. *arXiv preprint* arXiv:2111.00396; 2021.

[40] Gu A, Dao T, Ermon S, Rudra A, Ré C. HiPPO: Recurrent memory with optimal polynomial projections. *Advances in Neural Information Processing Systems*. 2020;33:1474–1487.

[41] Tay Y, Dehghani M, Abnar S, Shen Y, Bahri D, Pham P, Rao J, Yang L, Ruder S, Metzler D. Long Range Arena: A benchmark for efficient transformers. *arXiv preprint* arXiv:2011.04006; 2020.

[42] Vaswani A. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017.

[43] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint* arXiv:2312.00752; 2023.

[44] Beck M, Pöppel K, Spanring M, Auer A, Prudnikova O, Kopp M, Klambauer G, Brandstetter J, Hochreiter S. xLSTM: Extended long short-term memory. *arXiv preprint* arXiv:2405.04517; 2024.

[45] Dosovitskiy A. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv preprint* arXiv:2010.11929; 2020.

[46] Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang X. Vision Mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint* arXiv:2401.09417; 2024.

[47] Alkin B, Beck M, Pöppel K, Hochreiter S, Brandstetter J. Vision-LSTM: xLSTM as generic vision backbone. *arXiv preprint* arXiv:2406.04303; 2024.

Just Accepted

[48] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: A self-configuring method for deep learning–based biomedical image segmentation. *Nature Methods*. 2021;18(2):203–211.

[49] Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. In: *BrainLes 2018, MICCAI Workshop*. Springer; 2019:311–320.

[50] Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: *MICCAI BrainLesion Workshop*. Springer; 2021:272–284.

[51] Ma J, Li F, Wang B. U-Mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint* arXiv:2401.04722; 2024.

[52] Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth HR, Xu D. UNETR: Transformers for 3D medical image segmentation. In: *WACV 2022*. 2022:574–584.

[53] Chen T, Ding C, Zhu L, Xu T, Ji D, Wang Y, Zang Y, Li Z. xLSTM-UNet can be an effective 2D & 3D medical image segmentation backbone with Vision-LSTM (ViL) better than its Mamba counterpart. *arXiv preprint* arXiv:2407.01530; 2024.

[54] Zhang M, Wu Y, Zhang H, Qin Y, Zheng H, Tang W, Arnold C, Pei C, Yu P, Nan Y, et al. Multi-site, multi-domain airway tree modeling. *Medical Image Analysis*. 2023;90:102957.

[55] Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Medical Physics*. 2011;38(2):915–931.

[56] Shi P, Hu J, Yang Y, Gao Z, Liu W, Ma T. Centerline boundary Dice loss for vascular segmentation. In: *MICCAI 2024*. Springer; 2024:46–56.

[57] Sun F, Luo Z, Li S. Boundary difference over union loss for medical image segmentation. In: *MICCAI 2023*. Springer; 2023:292–301.

[58] Acebes C, Moustafa AH, Camara O, Galdran A. The centerline-cross entropy loss for vessel-like structure segmentation: Better topology consistency without sacrificing accuracy. In: *MICCAI 2024*. Springer; 2024:710–720.

Just Accepted