



OPEN

Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks

Antonio Garcia-Uceda^{1,2}✉, Raghavendra Selvan^{3,4}, Zaigham Saghir⁵, Harm A. W. M. Tiddens^{1,2} & Marleen de Bruijne^{1,3}✉

This paper presents a fully automatic and end-to-end optimised airway segmentation method for thoracic computed tomography, based on the U-Net architecture. We use a simple and low-memory 3D U-Net as backbone, which allows the method to process large 3D image patches, often comprising full lungs, in a single pass through the network. This makes the method simple, robust and efficient. We validated the proposed method on three datasets with very different characteristics and various airway abnormalities: (1) a dataset of pediatric patients including subjects with cystic fibrosis, (2) a subset of the Danish Lung Cancer Screening Trial, including subjects with chronic obstructive pulmonary disease, and (3) the EXACT'09 public dataset. We compared our method with other state-of-the-art airway segmentation methods, including relevant learning-based methods in the literature evaluated on the EXACT'09 data. We show that our method can extract highly complete airway trees with few false positive errors, on scans from both healthy and diseased subjects, and also that the method generalizes well across different datasets. On the EXACT'09 test set, our method achieved the second highest sensitivity score among all methods that reported good specificity.

Segmentation of the airway tree from thoracic Computed Tomography (CT) is a useful procedure to assess pulmonary diseases characterized by structural abnormalities of the airways, such as bronchiectasis (widening of the airways) and thickening of the airway wall. To quantitatively assess these conditions on the CT scan, clinicians are interested in having individual airway measurements, including airway diameter, wall thickness and tapering^{1,2}. The bronchial tree is a complex 3D structure, with many branches of varying sizes and different orientations. Segmenting airways by manual or semi-automatic methods is extremely tedious and time-consuming, taking more than 15 h¹ per scan, manually; or up to 2.5 h³ per scan, semi-automatically. Thus, a fully automatic airway segmentation method is important to provide an accurate, effortless and free of user-bias segmentation of the airway tree needed to quantify airway abnormalities.

In CT scans, airways appear as tubular structures with an interior of typically low intensity (the lumen) enclosed by a structure of higher intensity (the airway wall). Moreover, airways are surrounded by a background which can be of low intensity (the lung parenchyma) or higher intensity (typically the vessels). The earliest airway segmentation methods used the region growing algorithm^{4,5} to segment the airway lumen, relying on the intensity difference between the airway lumen and wall. Region growing can accurately capture the central bronchi, but has a tendency to systematically fail at extracting the peripheral bronchi of smaller size, missing a large portion of the airway tree. Also, when segmenting the smaller airways it often results in segmentations leaking into the lung parenchyma. This is due to the reduced intensity difference and reduced signal-to-noise ratio between the airway lumen and wall, caused by partial volume effects near the smaller airways. Many airway segmentation methods build upon the region growing algorithm, using this as an initial step to segment the larger bronchi and then apply additional operators to detect smaller airways while preventing leakage^{6–10}. Extensions to region growing based methods have been widely reviewed^{11,12}. In the airway extraction challenge EXACT'09¹³,

¹Department of Radiology and Nuclear Medicine, Erasmus MC, 3015 CE Rotterdam, The Netherlands. ²Department of Pediatric Pulmonology and Allergology, Erasmus MC-Sophia Children Hospital, 3015 CE Rotterdam, The Netherlands. ³Department of Computer Science, University of Copenhagen, 2100 Copenhagen, Denmark. ⁴Department of Neuroscience, University of Copenhagen, 2200 Copenhagen, Denmark. ⁵Department of Medicine, Section of Pulmonary Medicine, Herlev-Gentofte Hospital, Copenhagen University Hospital, 2900 Hellerup, Denmark. ✉email: a.garciauced@erasmusmc.nl; marleen.debruijne@erasmusmc.nl

15 algorithms were evaluated and compared, out of which 11 were region growing based methods. The results showed that all participating methods missed a significant amount of the smaller branches, and many methods had a large number of false positives errors.

Several airway segmentation methods use machine learning classifiers, either for voxelwise airway classification^{9,14} or to remove false positive airway candidates from a leaky baseline segmentation^{15,16}. These classifiers (kNN⁹, AdaBoost¹⁵, support vector machines¹⁶ or random forest¹⁴) operate on a set of predefined image features: multiscale Gaussian derivatives⁹, multiscale Hessian-based features^{14–16} or image texture features with local binary patterns¹⁴. These methods can achieve more complete airway tree predictions than previous purely intensity-based methods, with fewer false positives. However, they are highly dependent on the image features used to train the classifier, and may be time-consuming to apply due to the time needed to compute the image features^{9,14,16}. Recently, many state-of-the-art methods for medical image segmentation tasks have used deep learning¹⁷, and in particular convolutional neural networks (CNNs)¹⁸. The main advantage of deep CNN methods over classical learning-based techniques is that the extraction of relevant image features is done automatically from data in an end-to-end optimised setting. Several CNN-based methods have been applied for airway segmentation^{19–33}. Charbonnier et al.¹⁹ applied CNNs to detect and remove leakage voxels from a leaked region growing based segmentation. Yun et al.²⁰ applied the so-called 2.5D CNN, a pseudo 3D CNN which processes the three perpendicular 2D slices around each voxel, to perform voxelwise airway classification. The methods^{21–33} are based on the U-Net architecture³⁴. This model and its 3D extension^{35,36} are highly successful for medical image segmentation. The main advantage of the U-Net over voxelwise CNN models is that it can process entire images in one forward pass through its encoding/decoding structure, resulting in a segmentation map directly. Jin et al.²¹ and Meng et al.²² applied the U-Net on local volumes of interest around airways, being guided by the centerlines from a baseline segmentation²¹ or by tracking the airways extracted from the U-Net²². Garcia-Uceda et al.²³ applied the U-Net on large image patches extracted from the 3D scans, using various data augmentation techniques. Garcia-Uceda et al.³³ proposed a joint approach combining both 3D U-Net and graph neural networks (GNNs)³⁷. Qin et al.^{24,25} applied the U-Net for 26-fold prediction of the 26-neighbour connectivities of airway voxels, and segments airways by grouping voxels with at least one mutually predicted connectivity with a neighbour voxel. Zhao et al.²⁶ combined both 2D and 3D U-Nets with linear programming-based tracking to link disconnected components of the segmented airway tree. Wang et al.²⁷ used the U-Net with spatial recurrent convolutional layers and a radial distance loss to better segment tubular structures. Qin et al.^{28,29} extended the U-Net with feature recalibration and attention distillation modules that leverage the knowledge from intermediate feature maps of the network. Zheng et al.³⁰ proposed a General Union loss to alleviate the intra-class imbalance between large and small airways. Zhou et al.³¹ extended the U-Net with a multi-scale feature aggregation module based on dilated convolutions, to include more contextual information from a larger receptive field. Nadeem et al.³² used the U-Net followed by a freeze-and-grow propagation algorithm to iteratively increase the completeness of the segmented airway tree.

In this paper, we present a fully automatic and end-to-end optimised method to perform airway segmentation from chest CTs, based on the U-Net architecture. A preliminary version of this work was presented in²³. The proposed method processes larger 3D image patches, often covering an entire lung, in a single pass through the network. We achieve this by using a simple U-Net backbone with low memory footprint, and having efficient operations that feed image data to the network. This makes our method simple, robust and efficient. We performed a thorough validation of the proposed method on three datasets with very different characteristics, from subjects including pediatric patients and adults and with diverse airway abnormalities, including the EXACT'09 public dataset¹³. We compared the method with other state-of-the-art segmentation methods tested on the same data. On the EXACT'09 data, we compared with all the recent methods in the literature evaluated on these data, in terms of the average performance measures reported by the EXACT'09 evaluation. On the other datasets, we compared with the two methods previously evaluated on the same data^{9,10}. Moreover, we evaluated the accuracy of our method with respect to the presence of lung disease and airway abnormalities in the CT scans from the three datasets.

This article is organized as follows: in "Methods" section we explain the proposed methodology for airway segmentation, in "Data" section we present the data used for the experiments in this paper, in "Experiments" section we explain the experiments performed to validate the method, in "Results" section we present the results obtained, in "Discussion" section we discuss these results and explain the advantages and limitations of the proposed method, and in "Conclusions" section we give the conclusions for this paper.

Methods

Network architecture. The network consists of an encoder (downsampling) path followed by a decoder (upsampling) path, at different levels of resolution. Each level in the downsampling/upsampling path is composed of two $3 \times 3 \times 3$ convolutional layers, each followed by a rectified linear (ReLU) activation, and a $2 \times 2 \times 2$ pooling/upsample layer, respectively. After each pooling/upsample layer, the number of feature channels is doubled/halved, respectively. There are skip connections between opposing convolutional layers prior to pooling and after upsampling, at the same resolution level. The final layer is a $1 \times 1 \times 1$ convolutional layer followed by a sigmoid activation function. This network is schematically shown in Fig. 1.

For analysis of 3D chest scans, we found that the main bottleneck is the memory footprint of the network. To alleviate this, we use non-padded convolutions in the first 3 resolution levels of the U-Net, where the outmost layers of voxels in the feature maps are progressively removed after each convolution. We still use zero-padding in the remaining levels after the third, to prevent an excessive reduction of the output size of the network. This allows to reduce the memory footprint by approximately 30% compared to a model with zero-padding in all

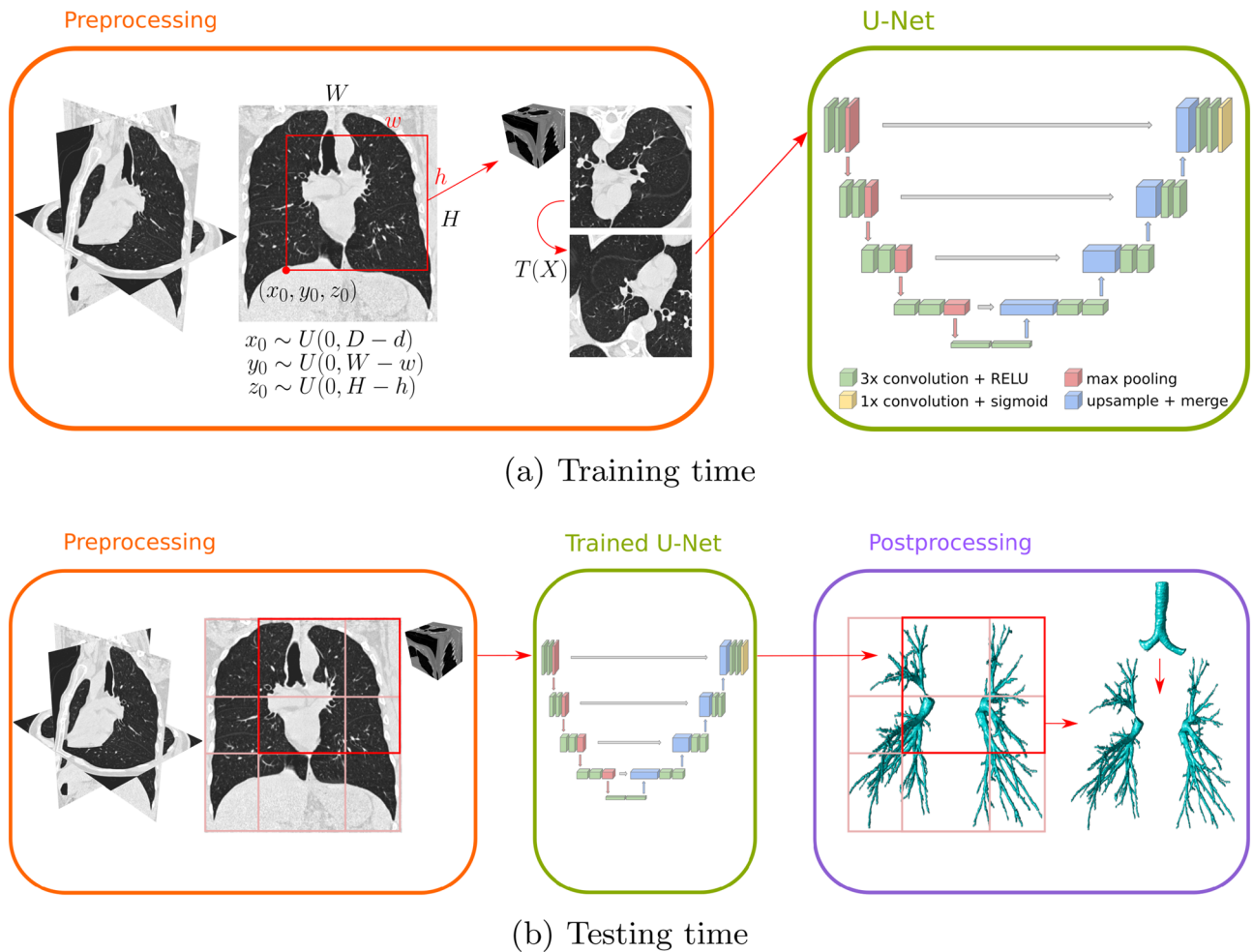


Figure 1. Schematics of the proposed airway segmentation method. **(a)** Overview at training time, showing the extraction of 3D random patches from the CT scan, to feed as input to the U-Net. **(b)** Overview at testing time, showing the extraction of 3D patches through sliding-window, and the generation of the airway tree segmentation from the patch-wise output of the U-Net.

layers. Moreover, we do not use dropout or batch normalization layers, as these require extra memory to store the feature maps after the operation.

Training of networks. The network is trained in an end-to-end supervised fashion, using the CT scans as input and evaluating the network output with respect to the ground truth, the reference airway segmentations, using the soft Dice loss³⁸,

$$\mathcal{L} = 1 - \frac{2 \sum_{x \in N_L} p(x)g(x)}{\sum_{x \in N_L} p(x) + \sum_{x \in N_L} g(x)} \quad (1)$$

where $p(x)$, $g(x)$ are the predicted voxelwise airway probability maps and airway ground truth, respectively. In the loss computation, the ground truth is masked to the region of interest (ROI): the lungs, indicated by the sub-index L , so that only voxels within this region contribute to the loss. This mask removes the trachea and part of the main bronchi from the ground truth, so that the training of the network focuses on the smaller branches. The lung segmentation needed for this masking operation is easily computed by a region growing method⁹.

Implementation of networks. When building the U-Net, the degrees of freedom (or hyperparameters) that determine the capacity of the network are (1) the number of resolution levels in the U-Net, (2) the number of feature maps in the first layer, (3) the input image size, and (4) the training batch size. We optimized these hyperparameters using subsets of training data from the two datasets used in this paper. We observed that the input image size was the most important parameter, and that by using larger input patches we obtained better results and faster convergence of the training and validation losses. With this, we found an optimal U-Net of 5 levels, 16 feature maps in the first layer, input size of $252 \times 252 \times 252$, and training batches containing only one image. This network can fit in a graphical card GPU NVIDIA Quadro P6000 with 24 GB memory during train-

ing. This U-Net construction is used for all the experiments undertaken in this paper. The implementation of the network is done using the Pytorch framework³⁹.

Generation of images during training. The 3D chest scans analysed in this paper have a size much larger than the input size of the network. Thus, we extract random patches of size $252 \times 252 \times 252$ from the full-size scans and ground truth to train the network. Moreover, the training datasets used for the experiments in this paper are small and contain only 15–25 scans. Thus, we apply data augmentation to the extracted image patches to increase artificially the amount of training data available and improve generalization. To generate input images to the network from the chest scans and ground truth, we apply the series of operations:

1. Crop the full-size scans to a bounding-box around the region of interest: the pre-segmented lung fields. This operation removes the areas outside the lungs, which are irrelevant for airway segmentation. The limits of the bounding-box are defined as 30 voxels from the outer voxel of the lung mask, in each direction. The extra 30-voxel buffer is used to prevent that boundary effects introduced by the use of zero-padding in the last layers of the network affect the prediction of peripheral airways, closer to the lung surface. Moreover, we mask the ground truth to the mask of the lung fields to remove the trachea and part of the main bronchi.
2. Extract image patches from the input image by cropping this to random bounding-boxes of $252 \times 252 \times 252$. We generate a total of 8 random patches per scan and per epoch during training. To perform this operation, we randomly sample the coordinates of the first corner of the bounding-box (x_0, y_0, z_0) from a uniform distribution in the ranges $x_0 \in [0, D - d]$, $y_0 \in [0, W - w]$ and $z_0 \in [0, H - h]$, respectively, where (d, w, h) is the size of the extracted patch, and (D, W, H) the size of the cropped input image. Then, the random bounding-box is generated with the limits $((x_0, x_0 + d), (y_0, y_0 + w), (z_0, z_0 + h))$.
3. Apply random rigid transformations to the input image patches as data augmentation. These transformations include (1) random flipping in the three directions, (2) random small 3D rotations up to 10 degrees, and (3) random scaling with a scale factor in the range (0.75, 1.25). The same transformations are applied to the ground truth using nearest-neighbour interpolation.

This series of operations is schematically shown in Fig. 1. We implemented operations (2) and (3) as an efficient on-the-fly image generation pipeline that feed data to the network, to reduce any computational overhead for the batch data loading part of the training algorithm. Operation (1) is applied only once to the chest scans and ground truth and prior to the training.

Airway extraction. To segment new scans, we extract image patches of size $252 \times 252 \times 252$ from the full-size scans in a sliding-window fashion in the three dimensions, with an overlap of 50% between the patches in each direction. Each image patch is processed by the trained network, producing a set of patch-wise airway probability maps. To obtain the full-size binary segmentation of the airway tree from these output patches, we apply the series of operations:

1. Aggregate the output patches by the network with airway probability maps by stitching them together (i.e. reversing the sliding-window operation used to extract the input patches). To account for the overlap between patches, we divide the aggregated voxelwise airway probabilities by the number of patches containing the given voxel.
2. Mask the full-size airway probability map to the mask of the lung fields. This is to discard the noise predictions outside the lung regions, as only these regions are included in the training of the networks by equation (1).
3. Threshold the airway probability maps to obtain the airway binary segmentation, using a threshold of 0.5 by default. This output does not contain the trachea and part of the main bronchi that were outside the mask of the lung fields. To include this, we merge the airway segmentation with a mask for the trachea and main bronchi. This is easily computed by a region growing method⁹, and then masking its output to the mask of the lung fields.
4. Apply connected component analysis⁴⁰ to the airway segmentation and retrieve the final airway tree as the largest 26-neighbour-connected component.

Data

To conduct the experiments, we used CT scans and reference airway segmentations from three different datasets:

1. **CF-CT:** This dataset consists of 24 inspiratory low-dose chest CT scans from pediatric patients at Erasmus MC-Sophia, 12 controls and 12 diseased¹. The 12 controls were patients with normal lung assessment from CT reported by two different radiologists, who received CT scanning for diagnostic purposes. The 12 diseased were 11 patients with Cystic fibrosis (CF) lung disease and 1 with Common Variable Immune Deficiency (CVID), who had structural lung damage with airway abnormalities visible on the CT scan. The two groups were gender and age matched, with the age range from 6 to 17 years old in both groups and 5 females per group. Scanning was undertaken using spirometry-guidance in a Siemens SOMATOM Definition Flash scanner. Each CT scan has an in-plane voxel size in the range 0.44–0.71 mm, with slice thickness in the range 0.75–1.0 mm, and slice spacing from 0.3 to 1.0 mm.
2. **DLCST:** This dataset consists of 32 inspiratory low-dose chest CT scans from the Danish Lung Cancer Screening Trial⁴¹. Participants to the trial were subjects between 50 and 70 years old, with a history of at least

- 20 pack-years of smoking, and without lung cancer related symptoms. Roughly half of the CT scans are from subjects with Chronic Obstructive Pulmonary Disease (COPD), reported from Spirometry measures. Scanning was undertaken using a MDCT scanner (16 rows Philips Mx 8000, Philips Medical Systems, Eindhoven, The Netherlands). All CT scans have a voxel resolution of $0.78 \times 0.78 \times 1 \text{ mm}^3$.
- EXACT'09:** This is the multi-center, public dataset of the EXACT'09 airway extraction challenge¹³. We used for evaluation purposes the EXACT'09 test set consisting of 20 chest CT scans. The conditions of the scanned patients vary largely, ranging from healthy volunteers to subjects with severe lung disease. The data includes both inspiratory and expiratory scans, ranging from clinical to low-dose. The CT scans were acquired with several different CT scanner brands and models, using a variety of scanning protocols and reconstruction parameters. Each CT scan has an in-plane voxel resolution in the range 0.55–0.78 mm, with varying slice thickness in the range 0.5–1.0 mm.

Generation of reference segmentations. Networks were trained and evaluated using airway lumen segmentations that were obtained by a combination of manual interaction and automated surface detection. For the CF-CT data, we have centerline annotations manually drawn by trained experts. For the DLCST data, we have airway extractions obtained from the union of the results of methods⁹ and¹⁰, and corrected by a trained observer. The visual assessment was done similarly to the EXACT'09 challenge¹³, where the obtained airways trees were divided in branches using a wave front propagation algorithm that detects bifurcations, and spurious branches were removed manually. “Partly wrong” branches whose segmentation covered the airway lumen but leaked outside the airway wall were also removed. To obtain the ground truth airway segmentations, we applied the optimal surface graph segmentation method⁴² on top of these initial airway references in order to (1) determine accurate lumen contours and (2) homogenize the ground truth for the two datasets. To evaluate the networks, we use these two ground truth segmentations as well as (via submission to the EXACT'09 challenge) the reference segmentations from the EXACT'09 test set. To build the EXACT'09 reference, the airway trees predicted by all 15 participating methods to the challenge were evaluated by independent expert observers, and the group of correct-scored branches were merged together.

Experiments

We performed three experiments to assess the performance of the proposed U-Net-based method in the different datasets and in comparison with previously published approaches:

- CF-CT:** We compared the performance of our method with that of a previously published airway segmentation method on these data⁴³. These results were obtained by a kNN-classifier together with a vessel similarity measure⁹, and refined with an optimal surface graph method⁴² to obtain an accurate lumen segmentation. We denote this method by **kNN-VS**. We trained and evaluated our U-Net model using the same 6-fold cross-validation setting as in⁴³. The training and testing data in each fold have an equilibrated number of scans from control and diseased subjects.
- DLCST:** We compared the performance of our method with that of two previously published airway segmentation methods on these data⁹ and¹⁰. The method⁹ is a kNN-classifier together with a vessel similarity measure, while the method¹⁰ extends the airways iteratively from an incomplete segmentation with locally optimal paths. We refined these results^{9,10} with an optimal surface graph method⁴² to obtain a more accurate lumen segmentation, and to homogenize them with the training data used to train our model, for a fair comparison. We denote these methods by **kNN-VS** and **LOP**, respectively. We trained and evaluated our U-Net model using the same 2-fold cross-validation setting as in⁹ and¹⁰. The training and testing data in each fold have a similar number of scans from healthy and diseased subjects.
- EXACT'09:** We compared the performance of our method with that of the 15 methods participating in the challenge EXACT'09 based on the results reported in¹³, 6 post-challenge methods evaluated on the EXACT'09 data and reported in⁴⁴, and 4 recent airway segmentation methods evaluated on these data^{19,20,29,45}. The methods in Charbonnier et al.¹⁹, Yun et al.²⁰ and Qin et al.²⁹ were previously described in the Introduction. The method in Gil et al.⁴⁵, named **PICASSO**, uses locally adaptive optimal thresholds learnt from a graph-encoded measure of the airway tree branching. We denote these methods by **CNN-Leak**, **2.5D CNN**, **U-Net-FRAD** and **PICASSO**, respectively. The method in Zheng et al.³⁰ is also evaluated on EXACT'09 data, but we did not include it in our comparison as the authors did not report the same specificity metric (false positive rate) as in the EXACT'09 evaluation. To our knowledge, these are all the recent methods in the literature that are evaluated on the EXACT'09 data. We trained our U-Net model using half of the CT scans from the CF-CT and DLCST data (28 scans in total). This training data has a similar number of scans from healthy and diseased subjects. We did not include the EXACT'09 training set in our training data as there are no reference segmentations available. Additionally, we compared our method with the nnU-Net segmentation method proposed by Isensee et al.⁴⁶, applied to airway segmentation. We denote this method by **nnU-Net**. To do this, we performed an experiment with the nnU-Net model: we trained it using the same training data as for our method, and we evaluated it on EXACT'09 data following the same evaluation protocol. We provide implementation details of the nnU-Net method and of our experiment in the Appendix.

For each cross-validation fold in the experiments, we used roughly 80% of images in the training fold to train the models, and the remaining 20% is used to monitor the validation loss. To train the models, we used the Adam optimizer⁴⁷ with an initial learning rate of 10^{-4} , which was chosen as large as possible while ensuring convergence of the training and validation losses. As convergence criterion, we monitored the moving average of the validation loss over 50 epochs, and training is stopped when its value (1) increases by more than 5%, or

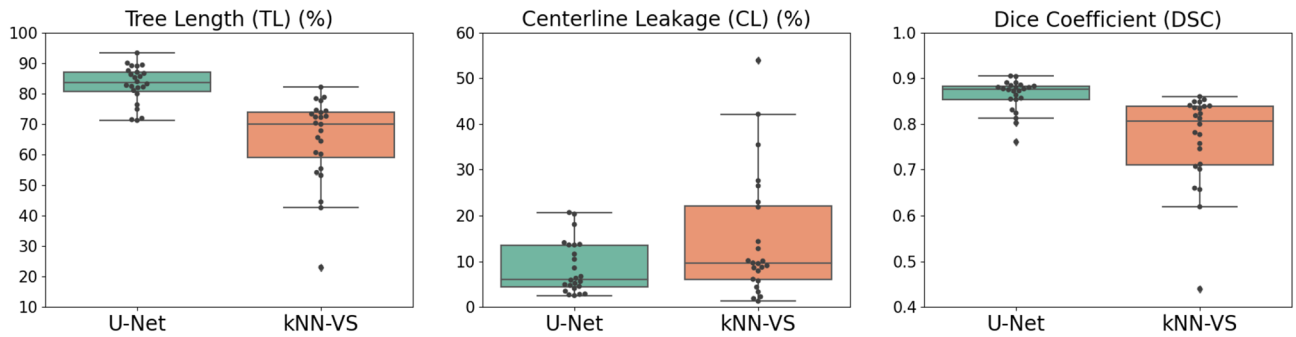


Figure 2. Boxplots showing (1) tree length detected, (2) centerline leakage and (3) Dice coefficient on the CF-CT data, for the results obtained by our U-Net-based method and the kNN-VS method⁴³. For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show all the data points.

(2) does not decrease more than 0.1%, over 20 epochs. We trained the models until convergence is reached, and we retrieved the model with the overall minimum validation loss for testing. Training time was approximately 1–2 days on a GPU NVIDIA Quadro P6000, while test time inference takes less than 1 min per scan including all post-processing steps to obtain the binary airway tree.

To compute the airway predictions on the EXACT’09 data, we thresholded the airway probability maps with a value of 0.1, and retrieved the final airway tree as the largest 6-neighbour-connected component. The lower threshold was to compensate for the reduction of completeness in our results when computing a single 6-connected structure, required by the submission guidelines for the evaluation on the EXACT’09 dataset¹³. In contrast, for the predictions on the CF-CT and DLCST data, we used the default threshold of 0.5 and 26-neighbour-connected component analysis.

Evaluation. To compare the different methods on the CF-CT and DLCST datasets, we computed (1) tree length detected, (2) centerline leakage and (3) Dice coefficient, with respect to the ground truth segmentations. For the results on the EXACT’09 data, we evaluated (1) tree length detected and (2) false positive rate, computed by the EXACT’09 challenge organizers¹³. Moreover, we assessed the segmentation accuracy of our U-Net-based method with respect to the presence of lung disease in the scans from each dataset. For this, we computed (1) tree length detected, (2) centerline leakage (or false positive rate on the EXACT’09 data) and (3) total tree length, and compared the measures from both healthy and diseased subjects.

The metrics are defined as:

1. **Tree length (TL):** The number of ground truth centerline voxels that are inside the predictions, as a percentage of the ground truth centerline full length.
2. **Centerline leakage (CL):** The number of predicted centerline voxels outside the ground truth, as a percentage of the ground truth centerline full length.
3. **False positive rate (FPR):** The number of false positive voxels outside the ground truth, as a percentage of the total number of ground truth voxels.
4. **Dice coefficient (DSC):** The voxelwise overlap between the predictions and the ground truth according to equation (2), with P and G the binary prediction and ground truth masks, respectively.

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad (2)$$

5. **Total tree length:** The sum of ground truth centerline voxels that are inside the predictions, multiplied by a factor that represents the voxel resolution (we used the geometrical mean of the voxel sizes in the three dimensions).

The trachea and main bronchi are removed in these measures from both the predictions and ground truth, similar to¹³. The centerlines are obtained by applying skeletonization⁴⁸ to the prediction and ground truth masks.

To compare the performance of the different methods, we used the paired two-sided Student’s T-test on the performance measures over the test set. To compare the accuracy of our U-Net-based method between healthy and diseased subjects, we used the unpaired two-sided Student’s T-test on the measures from the two groups. We consider that a p-value lower than 0.01 indicates that the two sets of measures compared are significantly different.

Results

Evaluation on CF-CT data. We show in Fig. 2 the evaluation on the CF-CT data of our U-Net-based method and the kNN-VS method⁴³. Comparing the U-Net with the kNN-VS results, the former shows higher TL (83.5(80.7–87.1) compared to 70.1(58.9–73.9), $p < 0.001$), higher DSC (0.876(0.854–0.883) compared to

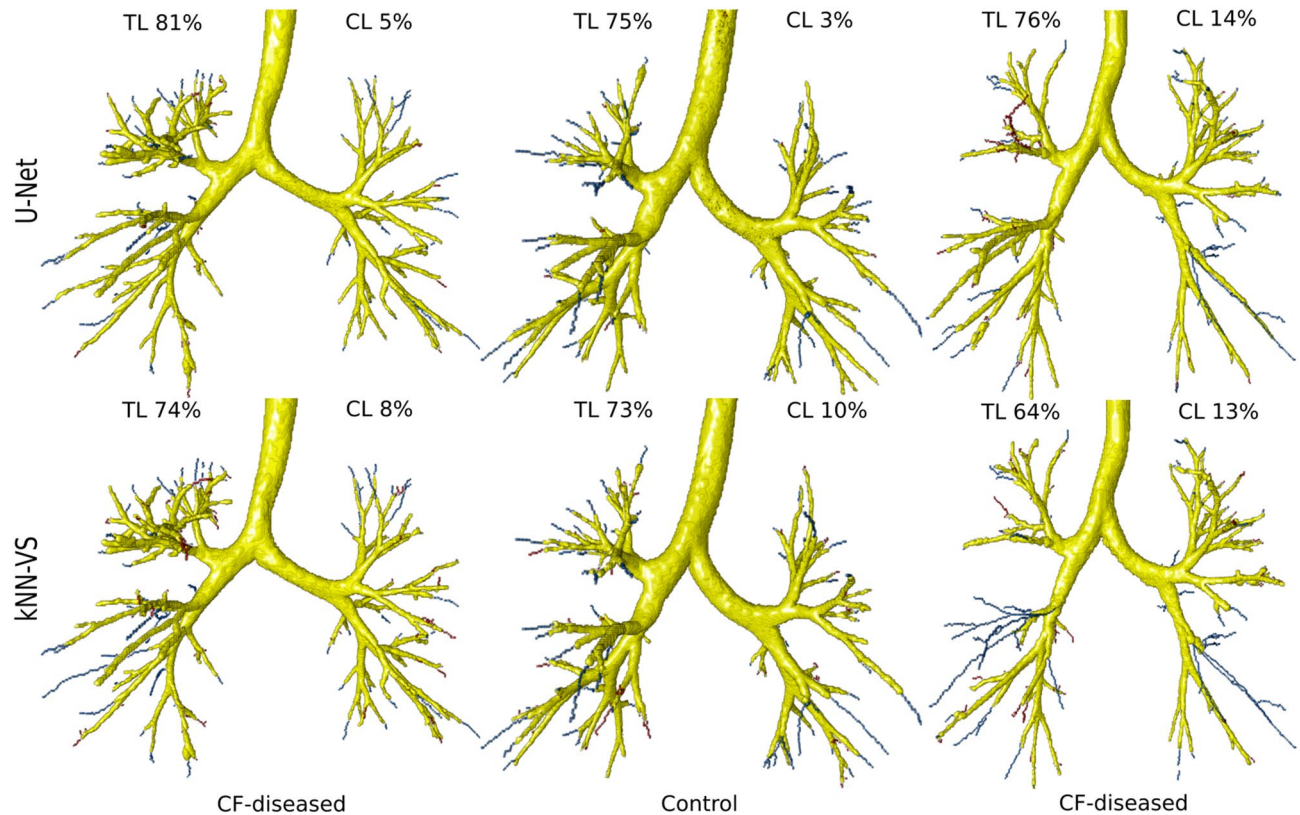


Figure 3. Visualization of predicted airways trees on the CF-CT data, obtained by our U-Net-based method and the kNN-VS method⁴³. True positives are displayed in yellow, centerline false negatives in blue and centerline false positives in red. These false negatives and positives are the errors counted in the tree length detected and centerline leakage measures reported in Fig. 2, respectively. The cases displayed correspond to, from left to right, the two U-Net results with Dice score closest to the median of the test Dice measures, and the U-Net result with the lowest Dice score.

0.806(0.711–0.839), $p < 0.001$) and a similar CL (6.09(4.41–13.6) compared to 9.58(5.98–22.1), $p = 0.021$). This indicates that our U-Net-based method predicts more complete airway trees than the kNN-VS method⁴³, with more and/or longer peripheral branches, and with a similar count of false positive errors. We show in Fig. 3 visualizations of airway trees obtained by the two methods. We can see that the airways predicted by our U-Net-based method match better the reference segmentations, with more and/or longer peripheral branches detected, while the kNN-VS method misses many of these peripheral branches.

Comparing the measures from our U-Net-based method between control and subjects with CF and airway abnormalities, we found no significant differences in TL (84.1(78.6–87.9) compared to 83.5(81.7–86.4), $p = 0.87$) and in CL (5.74(3.91–13.8) compared to 6.45(4.68–11.2), $p = 0.78$), but they were significant in total tree length (146(135–156) compared to 263(208–308), $p < 0.01$). This indicates that our method has similar accuracy with respect to the manual annotations on scans from both control and subjects with CF, while segmenting more airway branches in the diseased cases. We show these grouped measures in the Appendix.

Evaluation on DLCST data. We show in Fig. 4 the evaluation on the DLCST data of our U-Net-based method and the kNN-VS⁹ and LOP¹⁰ methods. Comparing the U-Net with the kNN-VS results, the former shows lower CL (8.25(6.26–9.58) compared to 12.0(9.68–14.9), $p < 0.001$), higher DSC (0.916(0.909–0.924) compared to 0.810(0.787–0.825), $p < 0.001$) and a similar TL (81.5(79.2–84.1) compared to 80.3(76.1–84.4), $p = 0.19$). Comparing the U-Net with the LOP results, the former shows lower CL (8.25(6.26–9.58) compared to 21.0(16.0–25.9), $p < 0.001$), higher DSC (0.916(0.909–0.924) compared to 0.795(0.780–0.812), $p < 0.001$), but lower TL (81.5(79.2–84.1) compared to 96.8(94.9–97.7), $p < 0.001$). This indicates that our U-Net-based method predicts an airway tree as complete as the kNN-VS method⁹, and less complete than the LOP method¹⁰, but with significantly less false positive errors than these two methods. It should be noted that this comparison is biased towards the kNN-VS and LOP methods that made up the reference segmentations, which partly explains the very high completeness of the results by the LOP method. We show in Fig. 5 visualizations of airway trees obtained by the three methods. We can see that the airways predicted by the LOP method have more peripheral branches detected (with almost no false negative errors), however they have more false positive errors. The results from our U-Net-based method have less peripheral branches detected, but also have much less false positives.

Through visual inspection of the results, we observed that several false positive errors correspond to real airways on the CT scan that were missing in the reference segmentations. This is because this reference was built

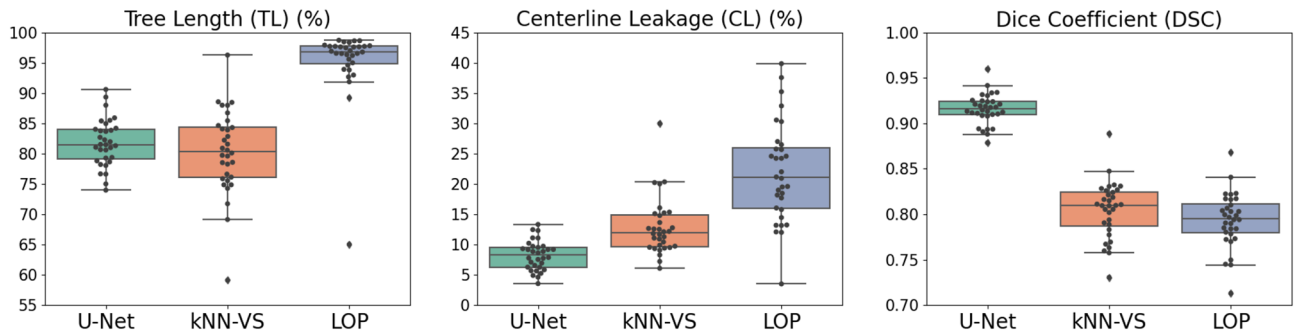


Figure 4. Boxplots showing (1) tree length detected, (2) centerline leakage and (3) Dice coefficient on the DLCST data, for the results obtained by our U-Net-based method and the kNN-VS⁹ and LOP¹⁰ methods. For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show all the data points.

in a conservative way from automatic airway extractions, where branches not extracted by either method^{9,10} were not included, and branches evaluated as “partly wrong” were removed, as explained in “[Generation of reference segmentations](#)” section. Thus, the reference segmentations are somewhat incomplete. For the segmentation results obtained by the three methods, we inspected any false positives that were of tubular shape and that were long enough to be clearly perceived as possible airways. We found 41 such cases of false positives that were actual airways when they were overlaid and analyzed on the CT scan. Out of these 41 misclassified branches, 20 were segmented only by our U-Net-based method, 14 were segmented by our U-Net-based method and either the kNN-VS or LOP method, and 7 were segmented by either the kNN-VS or LOP method and not the U-Net-based method. All cases segmented by our U-Net-based method were free of leakage. In contrast, all cases segmented by the kNN-VS and LOP methods had errors: they were branches longer than the real airway, or had leakage outside the airway wall. We show in Fig. 6 some examples of these misclassified false positive branches, where in the corresponding overlay of the airway mask with the CT scan we can see that they are real airways. Because of this, the CL leakage reported for the three methods in Fig. 4 are presumably overestimated, and to a larger extent for our U-Net-based method. Interestingly, most of the 34 misclassified but correctly segmented branches by our U-Net-based method were on CT scans from subjects with COPD: 26 out of 34 (76%).

Comparing the measures from our U-Net-based method between subjects with normal lung function and subjects with COPD, we found no significant differences in TL (81.0(78.8–81.5) compared to 82.7(80.0–85.2), $p = 0.15$), but they were significant in CL (6.27(5.25–8.86) compared to 9.18(7.64–10.4), $p < 0.01$) and in total tree length (190(174–247) compared to 131(123–169), $p < 0.01$). This indicates that our method has slightly more false positives errors on scans from subjects with COPD. However, this is partly explained by the higher number of misclassified but correctly segmented branches that we found in scans with COPD, for the U-Net results. We show these grouped measures in the Appendix.

Evaluation on EXACT’09 data. We show in Fig. 7 the evaluation on the EXACT’09 data of our U-Net-based method, the 15 methods from the challenge EXACT’09¹³, the 6 post-challenge methods⁴⁴, the 4 recent methods^{19,20,29,45} evaluated on the EXACT’09 data, and our experiment with the nnU-Net method⁴⁶ on these data. The reported results are the average TL and FPR measures obtained over the EXACT’09 test set. Our U-Net-based method achieves an overall TL of 70.3% and FPR of 2.74%. When compared to the EXACT’09 methods and the 6 post-challenge methods⁴⁴, our U-Net-based method has a TL much higher (at least 10%) than the scores from all methods except two, while the reported FPR is only slightly higher ($\approx 1\%$) than the average score among these methods. This indicates that our U-Net-based method predicts more complete airway trees on average, and with limited false positive errors. The two methods with higher TL also show a much higher FPR. When compared to the CNN-Leak method¹⁹ and the nnU-Net method⁴⁶, our U-Net-based method has higher completeness but also more false positive errors. The nnU-Net method⁴⁶ achieved a similar performance to the CNN-Leak method¹⁹, with slightly lower TL for a similar FPR score. When compared to the U-Net-FRAD method²⁹, our U-Net-based method has slightly lower completeness but also less false positive errors. Our method together with the CNN-Leak¹⁹, U-Net-FRAD²⁹ and nnU-Net⁴⁶ methods seem to have the best trade-off between the TL and FPR scores among all tested methods in Fig. 7. However, the authors from¹⁹ did not follow the independent evaluation for the EXACT’09 benchmark, but they re-implemented the evaluation algorithm using the reference standard from EXACT’09¹³. Their evaluation on one of the 15 submissions to EXACT’09 resulted in a slightly higher tree length than the one originally reported in¹³. Thus, the TL score reported for this method could be overestimated. Comparing our U-Net with the nnU-Net results, the former shows higher TL (68.8(61.2–79.7) compared to 63.9(52.8–75.2), $p = 0.01$) and higher FPR (1.90(0.51–3.86) compared to 1.07(0.30–2.25), $p = 0.04$), but they are not significantly different. This indicates that both our U-Net-based method and the nnU-Net have comparable accuracy. When compared to the 2.5D CNN method²⁰ and the PICASSO method⁴⁵, our U-Net-based method shows better performance measures, with higher completeness and less false positive errors.

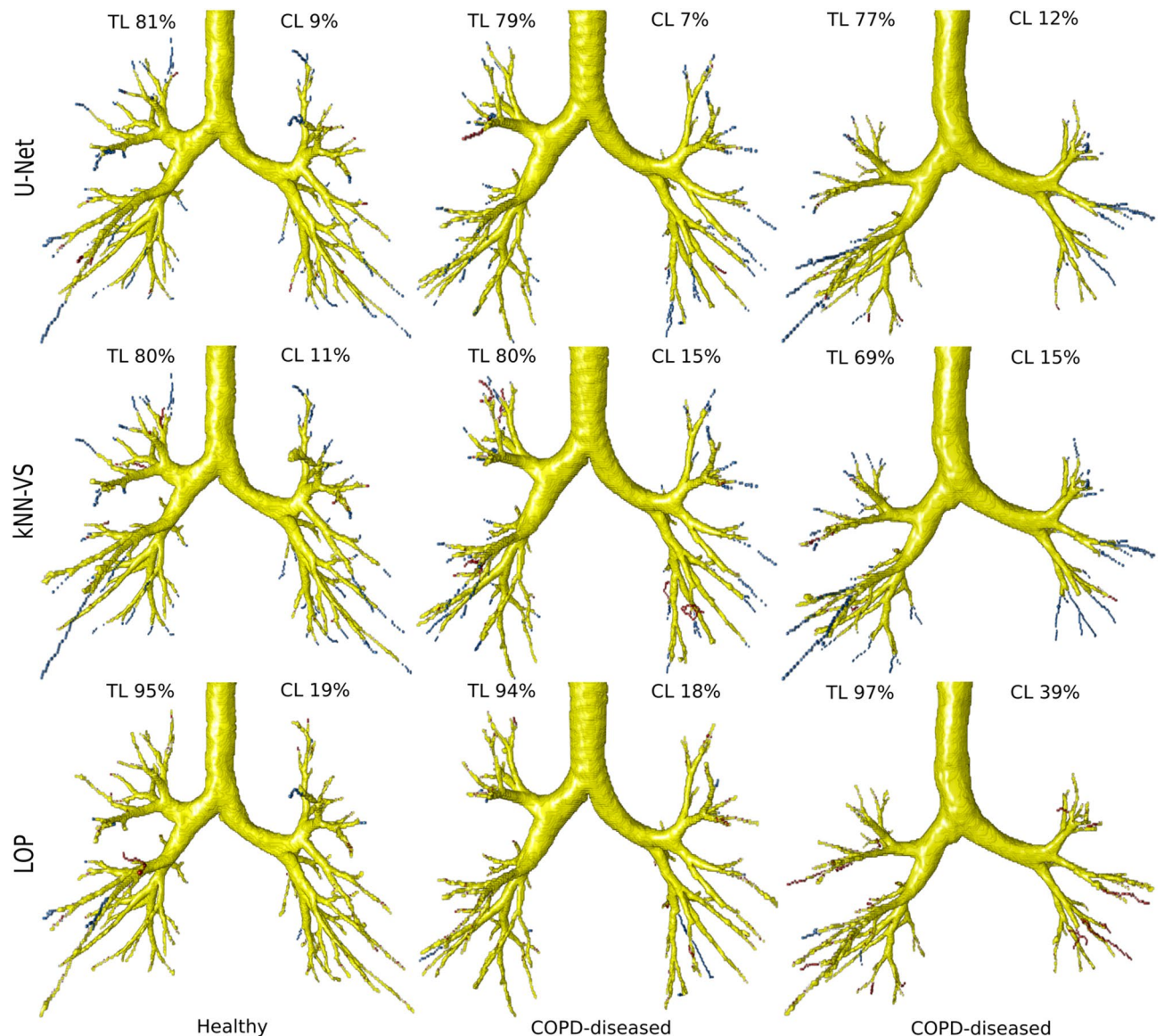


Figure 5. Visualization of predicted airways trees on the DLCST data, obtained by our U-Net-based method and the kNN-VS⁹ and LOP¹⁰ methods. True positives are displayed in yellow, centerline false negatives in blue and centerline false positives in red. These false negatives and positives are the errors counted in the tree length detected and centerline leakage measures reported in Fig. 4, respectively. The cases displayed correspond to, from left to right, the two U-Net results with Dice score closest to the median of the test Dice measures, and the U-Net result with the lowest Dice score.

Through visual inspection of the results, we observed that for some CT scans a large number of airways predicted by our U-Net-based method that were reported as false positives in the EXACT⁰⁹ evaluation are real airways on the CT scan. We show in Fig. 8 visualizations of airway trees, obtained by our method and the nnU-Net method⁴⁶, for which most of the reported false positive errors are real airways. Our method shows more misclassified false positive branches than the nnU-Net method, which partly explain the higher FPR scores in the displayed cases. This is because the reference standard from EXACT⁰⁹, constructed from all 15 participating methods to the challenge, is somewhat incomplete: it was estimated that on average approximately 10% more branches were visible in the scans than those that were included in the reference¹⁵. Thus, the false positive rate reported by our U-Net-based method and the nnU-Net method⁴⁶ in Fig. 7 may be overestimated, with a larger extent for our method. This could be also for the other methods^{19,20,29,44,45} we compared with, which used the same reference for evaluation.

We compared the measures from our U-Net-based method between a group of scans without any reported anomalies on the CT scan (cases 21, 22, 23, 28, 29, 35, and 37), and a group of scans with reported bronchiectasis (cases 33, 34, 36, and 39)¹³. We found no significant differences in TL (80.1(65.6–83.2) compared to 79.4(75.7–82.1), $p = 0.59$), in FPR (2.98(0.99–3.87) compared to 3.34(1.99–6.50), $p = 0.21$) and in total tree length (116(100–193) compared to 275(232–295), $p = 0.07$). This indicates that our method has similar accuracy

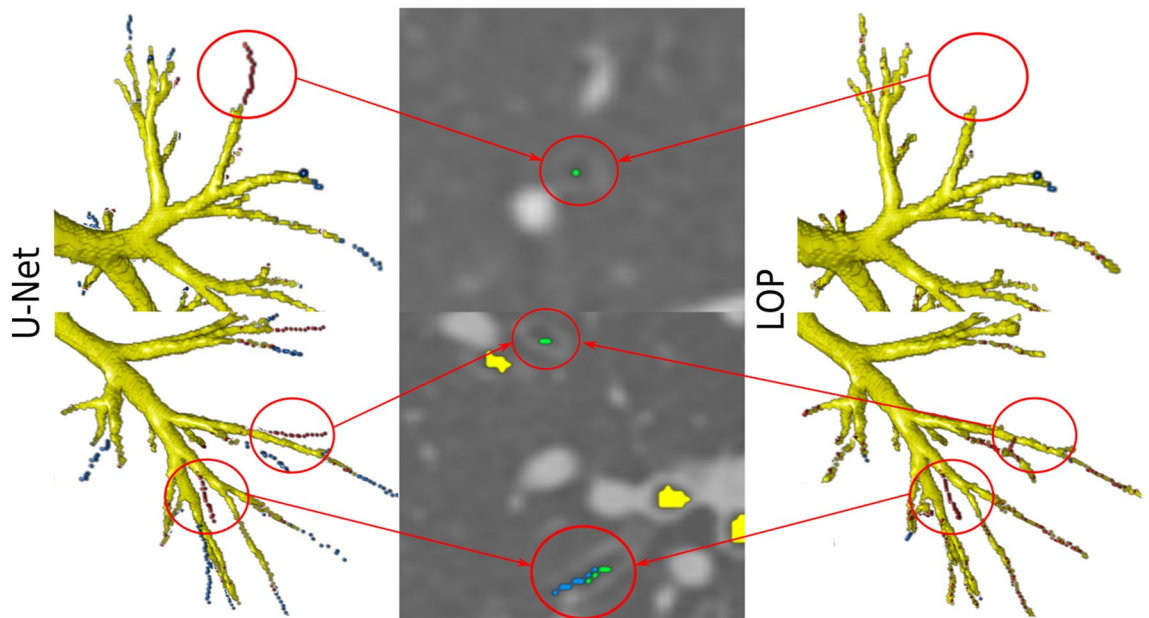


Figure 6. Visualization of examples of predicted false positive branches on the DLCST data, obtained by our U-Net-based method and missed by the kNN-VS and LOP methods, that are actually real airways missing in the reference segmentations. The last earmarked branch is detected by the LOP method but with errors: with leakage outside the airway wall. In the 3D visualizations the true positives are displayed in yellow, centerline false negatives in blue and centerline false positives in red. We also show the overlay of the predicted airway centerline masks with the CT scan, showing that the earmarked branches are real airways. The predicted centerlines by our U-Net method are displayed in green, those by the LOP method in blue, and the reference segmentations in yellow.

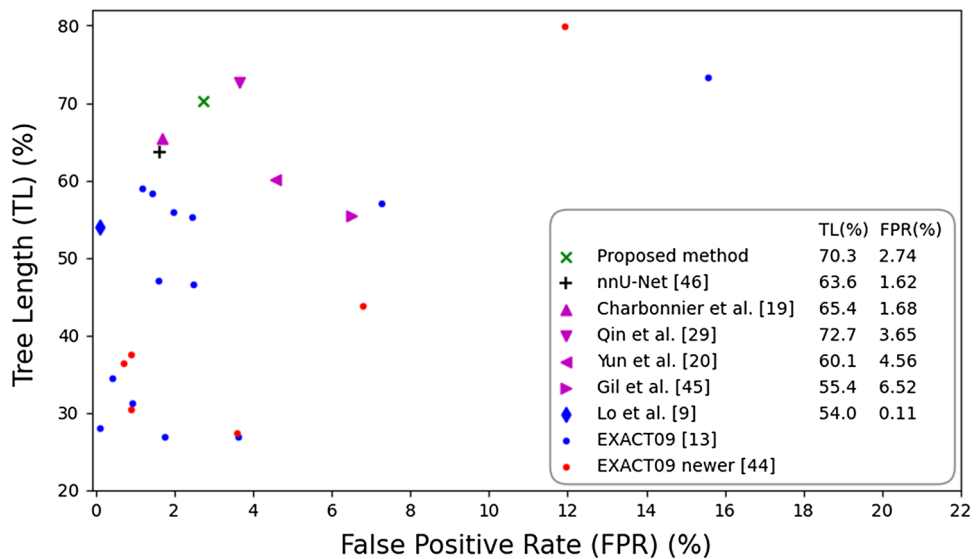


Figure 7. Average tree length detected versus average false positive rate over EXACT'09 test set, for our U-Net-based method, the 15 methods from the challenge EXACT'09¹³, the 6 post-challenge methods⁴⁴, the 4 recent methods^{19,20,29,45} evaluated on the EXACT'09 data, and our experiment with the nnU-Net method⁴⁶ on these data. Results are computed by the EXACT'09 challenge organizers except for¹⁹, who did the evaluation slightly differently, leading to slightly better performance, as reported in their paper.

with respect to the reference segmentations on both scans without anomalies and scans with bronchiectasis. We show these grouped measures in the Appendix.

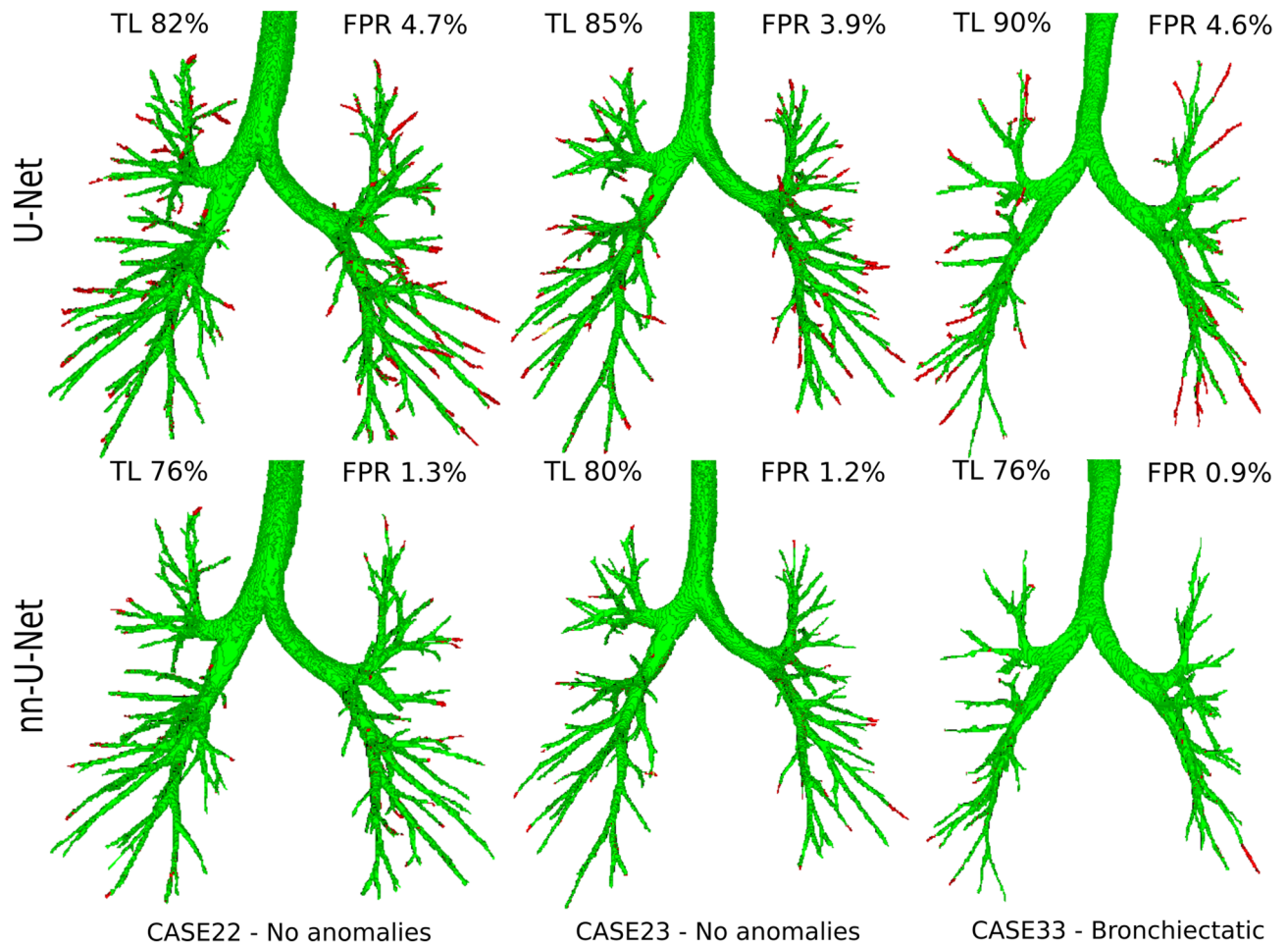


Figure 8. Visualization of predicted airway trees on the EXACT'09 data obtained by our U-Net-based method and the nnU-Net method⁴⁶. True positives are displayed in green and false positives in red, indicated according to the EXACT'09 evaluation. The cases displayed correspond to 3 out of the 5 U-Net results with the highest false positive rate reported in the EXACT'09 evaluation, where we found through visual inspection that many of the reported false positives are real airways on the CT scan, which were presumably missing in the EXACT'09 reference.

Discussion

Performance compared to other methods. The proposed U-Net-based method has shown good performance on the three different datasets, achieving highly complete airway segmentations with low false positive errors. The three datasets tested have very different characteristics: with subjects of a wide age range (from children in CF-CT data to older adults in DLCST data) and with airway abnormalities from diverse lung diseases, including CF and COPD. On the CF-CT and DLCST data, our U-Net-based method obtained more accurate airway segmentations than the other tested methods^{9,10,43}, shown by higher completeness and Dice scores altogether, and lower false positive errors. On the EXACT'09 data, our U-Net-based method achieved performance measures similar to the best performing previous methods, with the second highest sensitivity score among all methods that reported good specificity. Especially, it is noticeable that our method achieved this good performance on the highly heterogeneous EXACT'09 data with models trained on different and more homogeneous datasets (CF-CT and DLCST). This shows the robustness and the capacity to generalize across different data of the proposed method. Furthermore, our U-Net-based method obtained similar performance measures on scans from both healthy and diseased subjects, on the CF-CT and EXACT'09 data. On the DLCST data, our method had slightly more false positives errors on scans from subjects with COPD, where airway detection can be more challenging due to emphysema. However, on scans with COPD we reported a higher number of false positives in the U-Net results that were real airways on the CT scan, which partly explains the lower specificity in the diseased cases. On the CF-CT data, the method segmented more airway branches on scans from subjects with CF, probably due to the widening of peripheral airways due to CF-bronchiectasis, which makes them easier to detect on the CT scan.

When compared to other CNN-based methods evaluated on the EXACT'09 data, our U-Net-based method shows a performance similar to that of the CNN-Leak¹⁹, U-Net-FRAD²⁹ and nnU-Net⁴⁶ methods. These four methods seem to have the best trade-off between sensitivity and specificity among all tested methods in Fig. 7.

However, the U-Net-FRAD method²⁹ was trained on data that included the EXACT'09 training set, using their own manually-drawn reference segmentations on these data. Since the scans from the EXACT'09 training and test sets have similar characteristics and scanning parameters¹³, this gives an advantage to the U-Net-FRAD method²⁹ over our U-Net-based method, which was trained on different data. On the other hand, the CNN-Leak method¹⁹ requires a coarse airway segmentation as initialization to the CNN-based leakage removal algorithm, and thus the completeness of the results is limited by that of the initial segmentation. In contrast, our U-Net-based method segments the full airway tree directly. Moreover, CNN-Leak¹⁹ applied leakage removal to a series of 15 runs of the region growing based algorithm⁴⁹, with different parameters that control the extent of leakage, upon which the results were merged. In fact, the results from applying the leakage removal only to the baseline segmentation⁴⁹ had a much lower sensitivity (TL 51.8% in contrast to 65.4%) and a slightly higher specificity (FPR 1.01% in contrast to 1.68%). Similarly, an ensemble of U-Net results with different settings would likely have a slightly better performance, but this is more time consuming. When compared with the nnU-Net method⁴⁶, our U-Net-based method achieved slightly higher completeness and false positive errors, but we found that the differences were not statistically significant. When compared with the 2.5D CNN method²⁰, our U-Net-based method shows better performance with both higher sensitivity and specificity. The higher accuracy of our method can be because the 2.5D CNN method processes three perpendicular 2D slices around each voxel, while the 3D U-Net can better capture the 3D morphological information of airways.

Regarding computational efficiency, our U-Net-based method has a low execution time during inference of about 1 min per scan, including all post-processing steps to obtain a single connected binary airway tree. It should be noted that computation times of different methods cannot easily be compared due to differences in hardware, but the following may give a rough idea of the efficiency of our method compared to previous work. The execution times reported by other methods evaluated on the EXACT'09 data are: the kNN-VS method⁹, 55 min per scan, although it was run on a much older CPU and was not parallelized; the CNN-Leak method¹⁹, 3–5 min per scan; the U-Net-FRAD method²⁹, approximately 50 s per scan, although it does not include any post-processing steps; the 2.5D CNN method²⁰, 2–8 min per scan; and the PICASSO method⁴⁵, approximately 10 min per scan. Our experiments with the nnU-Net method⁴⁶ required 5–15 min per scan, although this includes some expensive pre-processing operations applied on the fly to the test images, such as resampling. Additionally, the U-Net method³² reports approximately 6.5 min per scan.

Advantages of the proposed method. The proposed U-Net-based method processes large 3D patches extracted from the CT scan in a single pass through the network. This makes our method simple, robust, and efficient at inference time, as only a few large patches are processed to segment a full scan. Although processing large input patches does not necessarily imply better performance, we observed that using large input patches resulted in better results and faster convergence of the training and validation losses. This is likely due to the more efficient use of the data during training. In contrast, when using smaller patches to train the model, this requires sampling patches at every voxel to equally process the same region of the CT scan comprised in a larger patch, which results in a high number of total patches. Moreover, since convolution operations are accelerated on the GPU, it is more efficient computationally to process a large patch at once through the network, rather than sequentially loading to GPU and processing smaller patches to include the same large region of the CT. In contrast to our approach, other U-Net-based methods in the literature either (1) apply a small U-Net locally around detected candidate airway locations, processing many image patches per scan^{21,22}, or (2) apply a model with larger memory footprint on smaller image patches in a sliding-window fashion^{24–32}. With our simple U-Net-based method and without further processing of the network output (except for computing the largest connected component) we obtained highly complete and accurate airway segmentations on the three datasets tested. On the EXACT'09 data, our method achieved a similar performance to the more complex U-Net-FRAD method²⁹, which used manually annotated EXACT'09 data for model training; and the nnU-Net method⁴⁶, using a similar U-Net as backbone. An additional advantage of processing few, large patches by our method is that fewer edge artifacts are introduced when tiling together the predicted output patches of the network. These artifacts typically occur where the tiled predicted patches meet in the full-size output (or where the patches overlap if this occurs), because the predictions from each patch can be slightly different due to border effects when using non-valid convolutions, and can cause discontinuities in the predicted airway mask.

Limitations. A limitation of our validation of the proposed method is that we did not evaluate it on CT scans with severe airway disease: the CF-CT dataset include subjects with moderate CF disease, the DLCST dataset include subjects with moderate COPD, and the EXACT'09 test data has various airways abnormalities but only one scan with reported “extensive bronchiectasis”¹³. Testing the method on severe cases would be important to assess its generalizability to tackle challenging airways with abnormally deformed shapes due to severe disease. Moreover, our evaluation of segmentation performance with respect to the presence of disease on the DLCST dataset may not show the whole picture. This is because the reference segmentations on these data were built in a conservative way from automatic airway extractions^{9,10}, and could be less complete for scans with severe emphysema.

A limitation of our U-Net-based method is that the prediction of the airway tree output by the U-Net is not guaranteed to form a connected tree structure. This could complicate the automatic extraction of airway biomarkers based on these segmentations as some methods assume a fully connected tree as input^{2,43,50}. The airway predictions obtained for this paper had typically some segmented peripheral airways disconnected from the main airway tree, and these were discarded when computing the largest connected component, which reduced the completeness of our airway segmentations. Alternatively, airway measurements techniques that do not rely on fully connected trees, such as⁵¹, could be used. Several methods have been proposed to extract more complete,

connected tree structures. The voxel-connectivity U-Net formulation²⁴ aims to improve connectivity in the airway prediction, resulting however in a model that is significantly more complex than our U-Net used. The linear programming-based tracking module on top of U-Net²⁶ attempts to link disconnected components of airways from the U-Net output. The mean-field networks and graph neural networks (GNNs)³⁷ emphasize the prediction of connected tree-like structures by phrasing the tree extraction problem as graph refinement starting from an over-connected input graph. The joint U-Net-GNN method³³ attempts to integrate this tree-like modelling in the U-Net prediction. However, none of these methods can guarantee fully connected airway tree predictions.

Finally, the memory footprint of our U-Net-based method could be further reduced, which would allow us to fit even larger images to the network, and possibly the entire CT scan. It may be possible to reduce the number of feature maps, especially in the decoding path of the U-Net, without decreasing much the performance. Also, using the partially reversible U-Net formulation⁵² in our method could largely reduce its memory footprint, by lowering the number of activation maps in the network stored in memory. However, this may result in an increase of training time (the authors from⁵² reported a 50% increase for their tested 5-level U-Net model, similar to our U-Net).

Conclusions

In this paper, we present a fully automatic and end-to-end optimised method to segment the airways from thoracic CT, based on the U-Net architecture. In contrast to previous U-Net methods for airway segmentation, the proposed method processes large 3D image patches often covering entire lungs. This is achieved by using a simple and low-memory 3D U-Net as backbone. This makes the method robust and efficient, which is important if the method is deployed in clinical software. Our method obtained highly complete and accurate airway segmentations on three very different datasets including CT scans with various airway and lung abnormalities. On the EXACT'09 test set, our method achieved the second highest sensitivity score among all methods that reported good specificity; and it outperformed previous methods on the other datasets.

Code availability

The source code for our implementation of the proposed method, and the trained model for the results obtained on the EXACT'09 test set, are available here: <https://github.com/antonioguj/bronchinet>.

Received: 30 March 2021; Accepted: 21 July 2021

Published online: 06 August 2021

References

- Kuo, W. *et al.* Diagnosis of bronchiectasis and airway wall thickening in children with cystic fibrosis: Objective airway-artery quantification. *Eur. Radiol.* **27**, 4680–4689 (2017).
- Kuo, W., Perez-Rovira, A., Tiddens, H., de Bruijne, M. & Study Group, N. C. C. Airway tapering. An objective image biomarker for bronchiectasis. *Eur. Radiol.* **30**, 2703–2711 (2020).
- Tschirren, J., Yavarna, T. & Reinhardt, J. Airway segmentation framework for clinical environments. In *Proceedings 2nd International Workshop Pulmonary Image Analysis* 227–238 (2009).
- Mori, K., Hasegawa, J., Toriwaki, J., Anno, H. & Katada, K. Recognition of bronchus in three-dimensional X-ray CT images with application to virtualized bronchoscopy system. In *Proceedings 13th International Conference on Pattern Recognition* 528–532 (1996).
- Sonka, M., Park, W. & Hoffman, E. Rule-based detection of intrathoracic airway trees. *IEEE Trans. Med. Imaging* **15**, 314–326 (1996).
- Fetita, C., Prêteux, F., Beigelman-Aubry, C. & Grenier, P. Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation. *IEEE Trans. Med. Imaging* **23**, 1353–1364 (2004).
- Fabijańska, A. Two-pass region growing algorithm for segmenting airway tree from MDCT chest scans. *Comput. Med. Imaging Graph.* **33**, 537–546 (2009).
- Graham, M., Gibbs, J., Cornish, D. & Higgins, W. Robust 3-D airway tree segmentation for image-guided peripheral bronchoscopy. *IEEE Trans. Med. Imaging* **29**, 982–997 (2010).
- Lo, P., Sporring, J., Ashraf, H., Pedersen, J. & de Bruijne, M. Vessel-guided airway tree segmentation: A voxel classification approach. *Med. Image Anal.* **14**, 527–538 (2010).
- Lo, P., Sporring, J., Pedersen, J. & de Bruijne, M. Airway tree extraction with locally optimal paths. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 51–58 (2009).
- Pu, J. *et al.* CT based computerized identification and analysis of human airways: A review. *Med. Phys.* **39**, 2603–2616 (2012).
- van Rikxoort, E. & van Ginneken, B. Automated segmentation of pulmonary structures in thoracic computed tomography scans: A review. *Phys. Med. Biol.* **58**, 187–220 (2013).
- Lo, P. *et al.* Extraction of airways from CT (EXACT'09). *IEEE Trans. Med. Imaging* **31**, 2093–2107 (2012).
- Bian, Z. *et al.* Small airway segmentation in thoracic computed tomography scans: A machine learning approach. *Phys. Med. Biol.* **63**, 155024 (2018).
- Inoue, T., Kitamura, Y., Li, Y. & Ito, W. Robust airway extraction based on machine learning and minimum spanning tree. *Int. Soc. Opt. Photonics (SPIE)* **8670**, 141–149 (2013).
- Meng, Q. *et al.* Automatic segmentation of airway tree based on local intensity filter and machine learning technique in 3D chest CT volume. *Int. J. Comput. Assist. Radiol. Surg.* **12**, 245–261 (2017).
- Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).
- Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3431–3440 (2015).
- Charbonnier, J. *et al.* Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Med. Image Anal.* **36**, 52–60 (2017).
- Yun, J. *et al.* Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *Med. Image Anal.* **51**, 13–20 (2019).
- Jin, D., Xu, Z., Harrison, A., George, K. & Mollura, D. 3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels. In *International Workshop on Machine Learning in Medical Imaging* 141–149 (2017).

22. Meng, Q. *et al.* Tracking and segmentation of the airways in chest CT using a fully convolutional network. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 198–207 (2017).
23. Garcia-Uceda Juarez, A., Tiddens, H. & de Bruijne, M. Automatic airway segmentation in chest CT using convolutional neural networks. In *Image Analysis for Moving Organ, Breast, and Thoracic Images* 238–250 (2018).
24. Qin, Y. *et al.* AirwayNet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 212–220 (2019).
25. Qin, Y. *et al.* AirwayNet-SE: A simple-yet-effective approach to improve airway segmentation using context scale fusion. In *IEEE International Symposium on Biomedical Imaging (ISBI)* 809–813 (2020).
26. Zhao, T. *et al.* Bronchus segmentation and classification by neural networks and linear programming. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 230–239 (2019).
27. Wang, C. *et al.* Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 348–356 (2019).
28. Qin, Y. *et al.* Learning bronchiole-sensitive airway segmentation CNNs by feature recalibration and attention distillation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 221–231 (2020).
29. Qin, Y. *et al.* Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT (2021). [arXiv:2012.05767](https://arxiv.org/abs/2012.05767).
30. Zheng, H. *et al.* Alleviating class-wise gradient imbalance for pulmonary airway segmentation (2020). [arXiv:2011.11952](https://arxiv.org/abs/2011.11952).
31. Zhou, K. *et al.* Automatic airway tree segmentation based on multi-scale context information. *Int. J. Comput. Assist. Radiol. Surg.* **16**, 219–230 (2021).
32. Nadeem, S. *et al.* A CT-based automated algorithm for airway segmentation using freeze-and-grow propagation and deep learning. *IEEE Trans. Med. Imaging* **40**, 405–418 (2021).
33. Garcia-Uceda Juarez, A., Selvan, R., Saghir, Z. & de Bruijne, M. A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs. In *International Workshop on Machine Learning in Medical Imaging* 583–591 (2019).
34. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 234–241 (2015).
35. Çiçek, Ö., Abdulkadir, A., Lienkamp, S., Brox, T. & Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 424–432 (2016).
36. Milletari, F., Navab, N. & Ahmadi, S. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Fourth International Conference on 3D Vision (3DV)* 565–571 (2016).
37. Selvan, R. *et al.* Graph refinement based airway extraction using mean-field networks and graph neural networks. *Med. Image Anal.* **64**, 101751 (2020).
38. Sudre, C., Li, W., Vercauteren, T., Ourselin, S. & Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 240–248 (2017).
39. Paszke, A. *et al.* PyTorch: An imperative style, high-performance deep learning library. *Adv. Neural Inform. Process. Syst.* **32**, 8026–8037 (2019).
40. Fiorio, C. & Gustedt, J. Two linear time union-find strategies for image processing. *Theor. Comput. Sci.* **154**, 165–181 (1996).
41. Pedersen, J. *et al.* The Danish randomized lung cancer CT screening trial overall design and results of the prevalence round. *J. Thorac. Oncol.* **4**, 608–614 (2009).
42. Petersen, J. *et al.* Optimal surface segmentation using flow lines to quantify airway abnormalities in chronic obstructive pulmonary disease. *Med. Image Anal.* **18**, 531–541 (2014).
43. Perez-Rovira, A., Kuo, W., Petersen, J., Tiddens, H. & de Bruijne, M. Automatic airway-artery analysis on lung CT to quantify airway wall thickening and bronchiectasis. *Med. Phys.* **43**, 5736–5744 (2016).
44. Extraction of Airways from CT 2009 (EXACT'09). http://image.diku.dk/exact/new_results.php.
45. Gil, D., Sanchez, C., Borrás, A., Diez-Ferrer, M. & Rosell, A. Segmentation of distal airways using structural analysis. *PLoS ONE* **14**, 1–17 (2019).
46. Isensee, F., Jaeger, P., Kohl, S., Petersen, J. & Maier-Hein, K. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
47. Kingma, D. & Ba, J. A method for stochastic optimization. ArXiv e-prints, Adam (2017). [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
48. Lee, T., Kashyap, R. & Chu, C. Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graph. Models Image Process.* **56**, 462–478 (1994).
49. Van Rikxoort, E., Baggerman, W. & Van Ginneken, B. Automatic segmentation of the airway tree from thoracic CT scans using a multithreshold approach. In *The Second International Workshop on Pulmonary Image Analysis* 341–349 (2009).
50. Petersen, J. *et al.* Effect of inspiration on airway dimensions measured in maximal inspiration CT images of subjects without airflow limitation. *Eur. Radiol.* **24**, 2319–2325 (2014).
51. Estepar, R. *et al.* Computational vascular morphometry for the assessment of pulmonary vascular disease based on scale-space particles. In *IEEE International Symposium on Biomedical Imaging (ISBI)* 1479–1482 (2012).
52. Brügger, R., Baumgartner, C. & Konukoglu, E. A partially reversible U-Net for memory-efficient volumetric image segmentation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI* 429–437 (2019).

Acknowledgements

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under Grant agreement n. 115721 resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in kind contribution.

Author contributions

A.G. did the literature research, conducted the experiments, analysed the results, and wrote the manuscript; R.S. contributed to the analysis of results and writing of the manuscript; Z.S. collected data, H.T. collected data, M.B. supervised the literature research, experiments and analysis of results, and contributed to writing of the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-95364-1>.

Correspondence and requests for materials should be addressed to A.G.-U. or M.d.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021