Title

A CT-Based Automated Algorithm for Airway Segmentation Using Freeze-and-Grow Propagation and Deep Learning.

Journal

IEEE Transactions on Medical Imaging, 40(1)

Authors

Nadeem, Syed
Hoffman, Eric
Sieren, Jessica
et al.

Publication Date

2021

DOI

10.1109/TMI.2020.3029013

Peer reviewed

HHS Public Access

Author manuscript

IEEE Trans Med Imaging. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

IEEE Trans Med Imaging. ; PP: . doi:10.1109/TMI.2020.3029013.

CT-based Automated Algorithms for Airway Segmentation using Freeze-and-Grow Propagation and Deep Learning

Syed Ahmed Nadeem,
Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA

Eric A. Hoffman,
Departments of Radiology, Medicine and Biomedical Engineering, University of Iowa, Iowa City, IA

Jessica C. Sieren,
Departments of Radiology and Biomedical Engineering, University of Iowa, Iowa City, IA

Alejandro P. Comellas,
Department of Medicine, University of Iowa, Iowa City, IA

Surya P. Bhatt,
Lung Health Center, University of Alabama at Birmingham, Birmingham, Alabama

Igor Z. Barjaktarevic,
Department of Medicine, Ronald Reagan UCLA Medical Center, Los Angeles, CA

Fereidoun Abtin,
Department of Radiological Sciences, Ronald Reagan UCLA Medical Center, Los Angeles, CA

Punam K. Saha [Fellow, IEEE]
Departments of Electrical and Computer Engineering and Radiology, University of Iowa. Iowa City. IA

Abstract

Chronic obstructive pulmonary disease (COPD) is a common lung disease, and quantitative CT-based bronchial phenotypes are of increasing interest as a means of exploring COPD sub-phenotypes, establishing disease progression and evaluating intervention outcomes. Reliable, fully automated and accurate segmentation of pulmonary airway trees is critical to such exploration. We present a novel approach of multi-parametric freeze-and-grow (FG) propagation which starts with a conservative segmentation parameter and captures finer details through iterative parameter relaxation. First, a CT intensity-based FG algorithm is developed and applied for airway tree segmentation. A more efficient version is produced using deep learning methods generating airway lumen likelihood maps from CT images, which are input to the FG algorithm. Both CT intensity- and deep learning-based algorithms are fully automated, and their performance, in terms of repeat scan reproducibility, accuracy, and leakages, is evaluated and compared with results from a method where segmentation results were reviewed and corrected with manual intervention. Both new algorithms show a reproducibility of 95% or higher for total lung capacity (TLC) repeat CT

syedahmed-nadeem@uiowa.edu.

scans. Experiments on TLC CT scans from different imaging sites at standard and low radiation dosages show that both new algorithms outperform the comparison dataset in terms of leakages and branch-level accuracy. Considering the performance and execution times, the deep learning-based FG algorithm is a fully automated option for large multi-site studies.

Index Terms—

Airway tree; COPD; CT; deep learning; freeze-and-grow; multi-parametric model; parameter relaxation; segmentation; tree-leakages

I. Introduction

Quantitative computed tomography (CT)-based bronchial characterizations of metrics such as airway lumen area, wall thickness, branching patterns, etc., along with parenchymal characterizations of emphysema and air trapping are in growing use in multi-center studies seeking to establish disease sub-phenotypes leading to improved understanding of disease etiology as well as intervention outcomes in chronic obstructive pulmonary disease (COPD) and asthma. Multi-center studies using such imaging approaches include COPDGene [1], SPIROMICS [2, 3], MESA Lung [4], CanCOLD [5] and SARP. [6] Through a combination of the use of CT imaging protocols selected to harmonize quantitative measures across sites, scanner manufacturers and scanner models, and through careful coaching of research subjects to a standardized full inspiratory (total lung capacity: TLC) and full expiratory (residual volume: RV) lung volume [2], objective quantitative assessments have led to a number of novel observations, including: 1) when the same anatomic airway segments are evaluated across subjects, airway walls, on average, are thinner in smokers with COPD compared with normal non-smokers [7]; 2) there are common airway branching variants in the over-all population which confer a greater risk of acquiring COPD in both the smoking and non-smoking population [8]; 3) metrics associated with quantitative CT assessment of the lung parenchyma and airways provide subject clusters in which distinct patterns of non-imaging derived phenotypes emerge [9], and 4) differences in the total number of airway generations segmented from the CT scan provide a measure of disease progression, consistent with observations from microscope-based assessment of lung specimens suggesting that peripheral airways disappear prior to emphysematous destruction of the parenchyma. [10] While segmentation and automatic segment labeling [11, 12] out to two generations beyond the segmental airways is an industry standard, some user verification and correction remains. Despite successful deployment [2] into studies involving thousands of subjects, there remains a desire for approaches which reliably extend airway segmentation to the peripheral segments and which reduce segmentation failures including leakages into low density parenchymal regions associated with parenchymal pathology.

Various segmentation methods using shape [13], boundary [14], region-growing [15, 16], or other models have been explored in different medical imaging applications. While shape- and boundary-based methods have unique strengths and are popular, region-growing strategies have been widely adopted for segmentation of anatomic tubular structures due to the high complexity in their shapes and branching patterns. [17] Several CT-based

algorithms for airway tree segmentation have been reported in the literature. [18–20] Different approaches have been adopted for airway segmentation, which include—(1) region growing and thresholding [11, 21–24]; (2) morphologic and geometric model-based approaches [25–27]; and (3) hybrid methods combining the previous two approaches. [28–30] Region growing methods rely on CT attenuation differences between airway lumen, airway wall, and surrounding lung parenchyma. However, due to limitations of spatial resolution and various imaging artifacts, CT attenuation differences become less pronounced at smaller bronchi causing segmentation leakages into the lung parenchyma, especially in the presence of parenchymal pathologies associated with decreased parenchymal density. Model-based methods rely on morphologic and geometric feature-based classification rules characterizing airway tree anatomy in CT images. However, it is difficult to manually select a set of features that holistically represents inter-subject and disease-related variabilities and models imaging artifacts as well as between-scanner variabilities.

Over the last several years, deep learning approaches have become popular in computer vision and image processing and have demonstrated unique strengths in extracting optimum knowledge embedded in data and put it into a process without requiring any application specific rules or process level design. [31–33] Following this trend, deep learning has also garnered attention in medical imaging including applications related to lung imaging. Charbonnier *et al.* [34] performed a post-segmentation step to classify leakages by using three orthogonal 2-D slices as input to a convolutional neural network (CNN). [35] Jin *et al.* [36] adopted a fully three-dimensional (3-D) approach using a 3-D U-Net [37, 38] in conjunction with incomplete sampling along the airway centerline to train a shallow (two pooling layers) but wide network to detect small airway tubular structures. Yun *et al.*, employed a 2.5-D CNN to improve a fully automated segmentation result by extending peripheral branches and removing false positives. [39] Juarez *et al.* presented a 3-D U-Net with elastic deformation-based data augmentation, but it was evaluated in terms of dice coefficient over only six images with numerous leakages at peripheral branches. [40]

In this paper, we present two automated CT-based airway tree segmentation algorithms and evaluate their performance in terms of reproducibility, accuracy, and leakages. First, a conventional image processing method of freeze-and-grow is developed, which starts with a conservative segmentation parameter and progressively captures finer details through iterative volume freezing and parameter relaxation. Second, a deep learning network is designed, trained, and validated to compute an airway lumen likelihood map from chest CT scans. The deep learning and freeze-and-grow methods are coupled by feeding the deep learning-based lumen likelihood map as an input to the freeze-and-grow algorithm. The new methods have been applied to total lung capacity (TLC) CT scans from ongoing large multi-site COPD-related studies, and their performance has been compared with existing manually corrected airway segmentations from multi-center trials.

II Methods

Although CT values at airway walls have empirically been found to be close to a value of −450HU, often the values decay to −800HU or less at distal sites; see Figure 1. Thus, a single parameter region growing model is not optimal for CT-based airway segmentation. In

the past, researchers have used space-varying parameter models. [11] Here, we present an iterative parameter-shift model that starts with a conservative parameter and gradually shifts to generous ones until convergence. We refer to this multi-parametric segmentation model as a freeze-and-grow (FG) algorithm. [41] A major advantage of the FG approach is that it does not rely on optimum selection of local parameters, which is always a challenging task and prone to errors causing leakages or early airway branch termination. In this section, first, we present a CT intensity-based airway segmentation algorithm using FG and support vector machine (SVM)-based post-pruning methods. [42] Later, we describe a deep learning (DL) method, based upon the FG results, for airway lumen classification and apply it to improve the efficiency and performance of FG and post-pruning methods. The DL method provided computational efficiency whereas the FG method is computationally intensive.

A. Freeze-and-Grow Algorithm

For CT intensity-based airway segmentation, the FG algorithm starts with the CT image as input and maintains two volume markings—(1) confident airway volume (CAV) representing validated airway volume and (2) forbidden volume (FV) around leakage-roots prohibiting connectivity paths to enter into leakage regions. FV is essential for iterative threshold relaxation. The algorithm consists of three steps—(1) initialization, (2) iterative parameter relaxation and segmentation volume update, and (3) termination.

At initiation, the algorithm selects a seed voxel inside the trachea using the method by Mori *et al.* [43]; let s be the seed voxel. Also, the algorithm begins with $CAV = \{s\}$ and an empty FV. Moreover, it starts with a conservative CT intensity threshold parameter of −1000HU for airway lumen region growing. During an iteration, the algorithm relaxes the method by incrementing the threshold parameter; grows airway lumen region from CAV as seeds over the entire image space excluding FV; and checks for a potential leakage. The algorithm continues incrementing the threshold parameter until a new leakage occurs. At the occurrence of a leakage, the algorithm enters a leakage correction phase including CAV and FV augmentation; and finally, it moves to the next iteration. The FG algorithm terminates when the threshold range is exhausted or FV covers the entire outer surface of CAV. The threshold range of [−1000 HU −700HU] has been used in this paper. Intermediate steps during an iteration of the FG algorithm are schematically illustrated in Figure 2.

1) Initial Lumen Segmentation and Centerline Computation—Let Z^3, where Z is the set of integers, denote the voxel grid or image space. During an iteration, at a threshold t, the airway tree volume V_t is computed using region growing from CAV over $Z^3 - FV$ and applying a fuzzy distance transform [44] (FDT)-based dilation for smoothing and hole-filling. FDT-based dilation was preferred over a binary method to account for partial voxel effects and avoid artificial merging of locally separate lumen regions; see [45].

During the FG algorithm, segmented airway tree volume monotonically expands with iterations. To comply with the monotonicity of volume change, the skeletonization method by Palágyi [46] was modified so that the curve skeleton S_t of $V_t \supset CAV$ is a superset of the skeleton S_{CAV} of CAV. Also, a new algorithm was developed for pruning spurious branches in a curve skeletal tree. The method is based on a voxel-wise predecessor-follower

relationship in a curve skeleton. [45] Specifically, it uses the depth of each skeletal subtree emanating from a junction voxel as compared to its local scale. Let p be a skeletal voxel immediately following a junction voxel. The skeletal subtree from p is spurious, if $depth(p)/scale(p) < \alpha_s$, where $depth(p)$ denotes the depth of the skeletal subtree from p, and $scale(p)$ is the FDT value at p. The value of the parameter $\alpha_s = 3$ was experimentally determined by analyzing the $depth(p)/scale(p)$ ratio of true branches. A spurious subtree from p is removed by deleting p and all following voxels. After removing spurious branches in S_t, the algorithm enters the leakage detection step.

2) Leakage Detection—Leakage detection is completed in two steps: (1) preliminary detection of potential leakages and (2) secondary screening for likely leakages. Potential leakages are identified by locating skeletal voxels with large change in associated airway lumen volume. These leakages are subjected to further topologic and geometric analysis for secondary screening of likely leakages, which are passed into a correction phase.

Let V denote an airway lumen volume and $S \supset V$ be a reference subset; intuitively, S represents a skeletal-like structure. For a voxel $p \in S$, the *associated volume* $\gamma(p)$ is the set of voxels $v \in V$ such that $\nexists q \in S$ closer to v than p. Let $\chi(p)$ be the set of voxels in the skeletal subtree from p; the *subtree volume* $\psi(p)$ is the union of associated volumes of all voxels in $\chi(p)$. Consider the confident lumen volume CAV, its skeleton S_{CAV} and the current lumen volume V_t at threshold t; note that $S_{CAV} \subset CAV \subset V_t$. For a skeletal voxel $p \in S_{CAV}$, let $\gamma_{cav}(p)$ and $\gamma_t(p)$ denote its associated volumes in CAV and V_t, respectively. A skeletal voxel p is a *potential leakage site* if the change in its associated volumes $|\gamma_t(p) - \gamma_{cav}(p)|$ exceeds a threshold $\alpha_{\Delta\gamma}$. We have used $\alpha_{\Delta\gamma} = 10$ mm^3 as it represents a potential spherical leakage of three-voxel radius, which is difficult to visually detect.

If no potential leakages are detected, the algorithm increases the threshold parameter from t to t+1 and repeats the lumen segmentation with the updated region growing parameter; this step is repeated until a potential leakage is found. Once a potential leakage voxel $p \in S_{CAV}$ is detected, the secondary screening for a likely leakage at p is initiated. During this step, we compute the length $l(p)$ of the skeletal subtree $\chi(p)$, which represents volume-weighted average geodesic length from p to terminal skeletal voxels. [41] A measure of *pseudo cross-sectional area* $\eta(p)$ is computed by dividing the subtree volume $|\psi(p)|$ by $l(p)$. Finally, the subtree emanating from p is upgraded to a likely leakage for correction if $\eta(p)$ is significantly greater than the expected airway lumen cross-sectional area at p, which is $\pi * \mathrm{FDT}^2(p)$. Specifically, a potential leakage at p is a likely leakage if

$$\eta(p)/\left(\pi * FDT^2(p)\right) > \alpha_\eta . \quad (1)$$

The value of the parameter $\alpha_\eta = 3$, used for experiments, was determined by analyzing the ratio $\eta(p) / (\pi * \mathrm{FDT}^2(p))$ at valid branches.

3) Leakage Correction—This step locates the root of a leakage and deletes the subtree volume from the leakage-root. The objective is to maximally recover valid airway branches passing the located likely leakage voxel. The leakage-root is located using a tree traversal approach starting from the likely leakage voxel and checking validity of individual branches.

Forward-branch-length, airway wall association, and subtree volume-to-linear outreach ratio, defined on V_t and its skeleton S_t, are used to define branch validity.

The *forward-branch-length* at a skeletal voxel $p \in S_t$ is the Euclidean length of the longest branch among the one containing p and its immediate child-branches. Airway wall association of a branch is computed by analyzing CT-intensity values around boundary voxels of associated lumen volume. Let p be a boundary voxel of associated lumen volume of a target branch and $B^+(p)$ denote the set of extended background neighbors of p, which are not radially farther than two voxels. [41] A lumen boundary voxel p is *surrounded by airway wall* if $f^+(p) - \mu(p) < 5$ 0 HU, where $f^+(p)$ is CT intensity of the brightest voxel in $B^+(p)$, and $\mu(p)$ is the mean airway wall intensity in the parent branch. Finally, the *airway wall association* of a branch is positive if two-thirds of its boundary voxels are surrounded by airway walls. *Linear outreach* of a skeletal subtree $\chi(p)$ is the Euclidean distance between p and the farthest terminal voxel in $\chi(p)$. *Subtree volume-to-linear outreach ratio* at p is the ratio of the subtree volume $|\psi(p)|$ and the linear outreach of $\chi(p)$. Finally, a skeletal branch is valid if (1) forward-branch-length is greater than half of the average airway branch length at the same generation as determined empirically; (2) airway wall association is positive; and (3) subtree volume-to-linear outreach ratio is less than three times the cross-section area of the immediate parent branch.

The leakage-root detection algorithm starts from a likely leakage voxel p and performs a recursive tree traversal checking validity of individual branches. If the branch containing p fails the validity test, the entire skeletal subtree $\chi(p)$, including p, is deleted; otherwise, it initiates a correction process for each of its children branches $p_1, p_2, \cdots$. The recursive correction process is continued until terminal branches are processed. Let S_t^* be the skeletal tree after deleting all leakages.

4) Segmentation Volume Augmentation—The confident airway lumen volume is updated as $CAU = \cup_{p \in S_t^*} |\gamma_t(p)|$, where $\gamma_t(p)$ is the associated volume of p in V_t with S_t as the reference skeleton; also, S_{CAV} is updated as S_t^*. New forbidden regions are computed around individual leakage-roots to arrest future leakage through the same leakage site; see [41] for details.

5) Final Leakage Removal—A final leakage removal step is applied to remove small leakages mostly occurring at terminal branches. Such leakages are missed by the rapid growth criterion used for initial leakage screening and slowly accumulate over iterations. To remove such leakages, a classifier-based approach was applied using topologic and geometric features at skeletal voxels. [41] A support vector machine (SVM) [42] classifier with a radial basis function was trained using manually labeled leakage and non-leakage skeletal voxels. The final leakage labelling was obtained using a neighborhood voting method on the SVM outcome in a top-down skeletal tree traversal.

B. Deep Learning-Based Freeze-and-Grow Algorithm

Deep learning (DL) is applied to compute a voxel-level airway lumen likelihood map from a CT image, which is input to an FG algorithm for airway tree segmentation; see Figure 3. A

modified 3-D U-Net classifier was developed with three pooling and three de-convolutional layers. High memory and computational complexity restrict the use of an entire CT image through the network; so smaller sub-regions of 64×64×64 voxels were used as samples. CT intensity values were truncated between −1200 HU and 200 HU and scaled between −1 and 1. Training sub-regions were selected using random sampling over the lung space. Data augmentation was avoided since it involves image transformation, introducing additional image blur and potential artifacts related to systematic feature perturbation. Kernels of size 3×3×3 were used at every convolutional layer except at the last layer, where a 1×1×1 kernel was used. At the highest resolution, 64 feature maps were used, and along the contracting path, the number of feature maps was doubled at each down- sampling step leading to 128, 256, and 512 feature maps at three lower resolutions. At each up-sampling step along the expansive path, the number of feature maps was halved after a concatenation with the feature map from the contracting path at the same resolution. To account for the sparsity and non-uniformity of airway structures in the training data, a weighted binary cross-entropy loss function was applied:

$$Loss = -\beta\Sigma_{p\in O}\log\hat{y}_p - (1-\beta)\Sigma_{p\in O^-}\log 1-\hat{y}_p, \quad (2)$$

where O and O^- are the object and background regions in a given training sample sub-region, β is the class balancing weight, $\hat{y}_p$ is the predicted value at a voxel p. [38]

At runtime, a trained 3-D U-Net was applied to compute a voxel-level airway lumen likelihood map from a chest CT image as follows—(1) generate a 32-voxel-spaced grid in the target CT image; (2) compute the airway lumen likelihood map over 64×64×64 sub-region around each grid location; and (3) compute airway lumen likelihood value at each voxel by averaging likelihood values from eight sub-regions containing that voxel. We adopted this new strategy of using overlapping sub-regions to reduce effects of sub-region bias adding checkerboard-type sub-region boundary artifacts and improve fidelity of the likelihood map. For DL-based FG algorithm, the initial likelihood threshold of 0.95 was applied, and the range of likelihood map of 0.95 to 0.35 was searched at a step size of 0.05 during FG iterations. The lower bound was empirically determined by running the algorithm over an extended range of likelihood values and selecting the lowest likelihood value that added a new airway branch.

1) Deep Learning-based Final Leakage Removal—A 3-D U-Net classifier was developed to remove small terminal leakages in DL and FG-based airway tree segmentations. The classifier was posed as a three-class segmentation problem—valid airway lumen, leakages, and background. CT intensity and initial segmentation label are fed as separate channels resulting in a 64×64×64×2 input sample. Similar network architecture and data normalization, described in the previous section, were adopted. Weighted categorical cross-entropy loss function was used for training and validation, where the weight for a class was computed over all training samples as one minus the average volume fraction of that class. Training samples were generated by randomly sampling 64×64×64 sub-regions along the terminal, and two-generation ancestral branches. Sub-regions along the centerline of larger branches were ignored because no leakages were observed at larger branches using the DL and FG-based segmentation.

At runtime, sub-regions were sampled on terminal, and two-generation ancestral branches with 32-voxel geodesic distance between two nearest samples. At each voxel, the leakage likelihood value was computed by averaging likelihood values from all sub-regions containing that voxel, and it was deleted if the leakage likelihood value exceeded a preset threshold α_{leak}. For our experiments, we used $\alpha_{leak} = 0.7$, which was determined empirically as mean-plus-three-sigma of observed leakage likelihood values over leakages in the training dataset.

C. Experiments

Experiments were designed to examine the reproducibility of our methods and evaluate their accuracy at standard and low radiation CT imaging from different study centers. The accuracy of our methods was compared with datasets from the SPIROMICS project [1,2] where the automated airway segmentation results underwent an additional manual editing process following a strict quality control protocol. All airway paths were assured to reach at least the segmental level. The 5 main paths (RB1, RB4, RB10, LB1, and LB10) were assured to reach at least the segmental+2 generation. All obvious leaks were removed. The manually edited SPIROMICS airway masks are referred to below as “industry-standard segmentation after manual editing” and the details of imaging and image assessment are provided in Sieren *et al.* [2] Different airway segmentation methods are abbreviated as follows—(1) CT intensity-based freeze-and-grow (IN-FG), (2) deep learning-based freeze-and-grow (DL-FG), (3) industry-standard segmentation before manual editing (IS-BE), and (4) industry-standard segmentation after manual editing (IS-AE). Our experiments involved (1) human chest CT data, (2) training and validation, (3) reference segmentation masks and analysis.

1) Human Chest CT Data—Three different datasets of total lung capacity (TLC) chest CT scans used for our experiments.

$\text{Data}_{\text{stand}}$: TLC chest CT scans of 120 subjects (age (years): [42 80] ([Min Max]), 62.5±9.9 (mean±std.); 60 female, smoking (pack-year): 43.4±25.6; COPD status (N): GOLD 0 (38), GOLD 1 (20), GOLD 2 (25), GOLD 3 (22), GOLD 4 (15)) were randomly selected from baseline visits of the subpopulations and intermediate outcome measures in COPD study (SPIROMICS). [3] This dataset includes scans from ten different sites with a variety of scanners utilizing scan protocols outlined in Sieren *et al.* [2] This dataset was used for training and validation of different algorithms as well as performance evaluation in terms of accuracy.

$\text{Data}_{\text{repro}}$: A set of forty pneumotachometer-monitored [47] TLC chest CT scans were acquired on a Siemens Sensation 64 (N = 15) or Siemens SOMATOM Definition Flash 128 (N = 5) scanner (Forchheim, Germany) from twenty subjects (age (years): [21 64], 29.6±11.1; 10 females; 8 current/former smokers), where each subject underwent two repeat scans at full inspiration 3 to 5 minutes apart. [48] Between scans, the subjects got off the scanner table, walked around, and returned to the table. This data set was used to examine reproducibility of the new methods.

$\underline{\text{Data}_{\text{low}}}$: Low radiation TLC chest CT scans of twelve non-smoking normal subjects (age (years): [20 54], 35.4±12.0; 5 females; all non-smokers). These scans were acquired on a Siemens SOMATOM Force scanner using 120 kV, LD CareDose reference 36 mA, 512×512 image matrix, 0.6 mm slice thickness, 0.5 mm slice spacing, pitch of 1.0, and low radiation dose of 1.3 mSv versus 3.2 mSv standard dose. This data set was used to examine accuracy of different methods at low radiation CT imaging.

2) Training and Validation—Twenty scans were randomly selected from $\text{Data}_{\text{stand}}$ and used for training and validation of (1) SVM leakage classifier, (2) DL-based airway lumen classifier, and (3) DL leakage classifier. Among these scans, fifteen were used for training and five for validation. For SVM leakage classifier, 6,489 leakage and 87,370 non-leakage samples were used for training, while 1,922 leakage and 22,986 non-leakage samples were used for validation. A leakage sample was collected at a skeletal voxel with leakages in its associated volumes, while non-leakage samples were obtained at other skeletal voxels. All training samples for SVM leakage classifier were collected at terminal branches and their 2-generation ancestors.

For DL airway lumen classifier, manually edited airway lumen segmentation using IS-AE were used for training and validation. For each scan, 2000 sub-regions of size 64×64×64 were randomly sampled yielding a set of 30,000 training and 10,000 validation samples. The network training step converged after twenty epochs using Adam optimization algorithm [49] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and a learning rate of 1×10^{-4}. The network training process took 40 hours running on an Intel Core i9–79000X CPU using a NVIDIA GeForce GTX 1080 Ti graphics card. For DL leakage classifier, 250 64×64×64 sub-regions with leakages and another 250 sub-regions without leakages were sampled at centerline locations from terminal branches and their 2-generation ancestors from each scan resulting 7,500 training and 2,500 validation samples. The training process converged after thirteen epochs using Adam optimization algorithm with β_1=0.9, β_2=0.999 and a learning rate of 1×10^{-4} taking 6 hours and 18 minutes.

3) Reference Segmentation Mask and Analysis—Reference segmentation masks (again utilizing data generated via SPIROMICS [2]) were used to assess accuracy. Considering the difficulties with manual outlining of true airway masks, a method was designed to generate unbiased reference airway masks. For a given scan, the segmentation masks generated using each of the three methods, namely, IS-AE, IN-FG, and DL-FG, were manually corrected by an expert. Finally, a reference airway mask was computed as the union of these manually corrected volumes.

Accuracy of an airway segmentation method was examined as the percentage of true branch detection along the five segmental airway paths (RB1, RB4, RB10, LB1, LB10) up to two generations beyond. Also, the branch- level accuracy along all airway paths at different generations was analyzed. Moreover, the residual leakage performance was examined in terms of leakage count and total leakage volume. Reproducibility was examined along the five standardized segmental paths and at different airway tree generations. There was no data overlap between training and testing datasets. Thus, one hundred datasets from $\text{Data}_{\text{stand}}$ were used for accuracy analysis at standard dose CT imaging. Note that $\text{Data}_{\text{repro}}$ and

$Data_{low}$ were not used for training; thus, they were used in their entirety for performance analysis.

III. Results and Discussion

Results of DL-based voxel-level airway lumen likelihood classification are presented in Figure 4 using color-coded maximum intensity projections (MIPs). Checkerboard type sub-region boundary artifacts are visible in (a), where 64 voxel sample grid-spacing without overlaps were used. Also, uncertain regions (greenish yellow) appear in (a) around small to medium airway branches. Both checkerboard artifacts as well as uncertain regions mostly disappear in (b), when a sample grid spacing of 32 voxels with 32-voxel overlapping was used. Further increase in sample density with grid spacing of 16 voxels significantly increased computation time to 854 seconds from 149 seconds for 32-voxel grid spacing without noticeable improvements in lumen classification; see (c). It is notable that likelihood averaging from adjacent sub-regions adds no lumen blurring even at fine scales; see (b,c). Finally, while the lumen structures are visible in the MIPs of (a-c), lumen color grades at finer scales are different justifying the use of multi-parametric FG algorithm for final segmentation.

Intermediate results of the DL-FG algorithm at different FG iterations are shown in Figure 5. DL-FG adds new airway branches at each iteration and completes the entire tree segmentation in thirteen iterations. For the same scan, IN-FG took 95 iterations to complete the airway tree segmentation producing visually similar results as DL-FG. It may be clarified that the result after the first iteration is the optimum segmentation result using a single threshold parameter on DL-based airway lumen classification, and new branches at subsequent iterations are added through the FG framework.

A Reproducibility

Figure 6 illustrates repeat scan reproducibility of airway tree segmentation using the DL-FG algorithm. For the subject on the top, segmented airway trees in repeat scans have fewer branches as compared to the one on the bottom. For both subjects, repeat scan reproducibility of airway tree segmentations are visually satisfactory with only a few minor differences at peripheral branches. Quantitative results of repeat scan reproducibility on $Data_{repro}$ are summarized in Figure 7. Mean and variance of branch-counts at different generations are shown in (a). Branch-count variability up to 5^{th} generation is ‘0’ using both methods, while peak branch- counts for both methods occur at 8^{th} generation. Beyond 8^{th} generation, average branch-counts for both methods continue decaying suggesting increased numbers of missing branches. Agreement of segmented branches in repeat scans are shown in (b). Both IN-FG and DL-FG algorithms deliver 97.7% or higher repeat scan reproducibility up to 8^{th} generation where the peak mean branch-count occurs in (a).

For the five segmental bronchial paths (RB1, RB4, RB10, LB1, and LB10) and two generations beyond, both algorithms were found 100% reproducible except at the sub-sub-segmental level along RB4. Both algorithms successfully detected the five segmental paths and their expected sub-segmental branches in both repeat scans. At the sub-sub-segmental level, both algorithms detected expected four branches along each segmental path in both

repeat scans, except along RB4, where, on an average, IN-FG and DL-FG algorithms detected 3.6 and 3.5 branches, respectively, with 98.6 and 97.2% reproducibility.

B. Accuracy at Standard CT Dosage

Results of accuracy experiments at standard CT radiation dosage for different algorithms are presented in Figures 8 and 9 and Table 1. Figure 8 presents the reference mask and airway tree segmentation results for two subjects. For the subject at the top row, IS-BE produced two leakages near LB1 and a terminal branch in the upper-right lobe, which were removed by IS-AE; also, a few branches were added; see the upper-left and right lobes. Results using IN-FG and DL-FG are visually similar to the reference segmentation mask with no apparent leakages. The two new methods captured additional branches near LB1 missed by the industry-standard method even after manual editing. For the case at the bottom row, there are no apparent leakages, and more branches are captured. Among different methods, both IN-FG and DL-FG captured more branches than the industry-standard methods. An entire LB1 subtree was added by both IN-FG and DL-FG, which were missed by IS-BE and partially added by editing (IS-AE). Also, the industry-standard methods missed numerous peripheral branches, which were automatically captured by the two new methods. Between the two new methods, there are only a few minor differences at peripheral branches, e.g., at the top row near RB10. Figure 9 presents results of accuracy analysis. Mean and variability of branch-counts in reference masks are shown in (a), while the branch-level accuracies are presented in (b). IS-AE and the two new methods captured all branches up to the 5th generation covering all segmental bronchi. At the 9^{th} generation, where the peak branch count occurs in the reference masks, IS-BE, IS-AE, IN-FG, and DL-FG captured 87.0, 90.4, 98.4, and 95.4% of true branches, respectively. The accuracy results beyond 12^{th} generation are insignificant as the number of branches captured beyond this generation are fewer.

For 100 subjects in $Data_{stand}$, IS-AE as well as the two new methods captured all five segmental bronchi for all subjects except for two cases—one missing LB10 and another missing RB10. IS-BE missed three additional segmental branches, which were manually added (IS-AE). All algorithms including IN-FG missed two expected sub-segmental branches for one subject. Both DL-FG and IS-AE missed one additional sub-segmental branch, while IS-BE missed another 17 sub-segmental branches. At the sub-sub-segmental level, IN-FG missed 44 branches, while DL-FG missed 42 branches. At this level, the industry-standard method missed 260 and 56 branches before and after editing. In summary, the branch level performance of the three methods IN-FG, DL-FG, and IS-AE are similar with minor improvements using the new fully automated methods, especially at the sub-sub-segmental level.

Results of leakage analysis are summarized in Table 1. Prior to editing, on an average, the industry-standard method generated 7.4 leakages at a volume of 449 mm^3 per image, which were reduced to 1.2 leakages at 54 mm^3. Both IN-FG and DL-FG produced fewer than one leakage at significantly reduced leakage volume per image (Table 1). The largest leakage observed for DL-FG was 29 mm^3 or approximately 5×5×5 voxels. Also, the farthest leakage voxel for DL-FG and IN-FG among all 100 cases examined in this experiment were 6 mm and 9 mm, respectively. Thus, the leakages identified for the two new methods are borderline

cases and are within the regime of lumen delineation variability. The largest leakage in the industry-standard method was 7,571 and 1,642 mm^3 with farthest voxel at 40 and 15 mm before and after manual editing, respectively. Also, the high standard deviation in total leakage volume per image using the industry-standard methods suggest high performance unreliability. Notably, the leakage performance variability is significantly low for the new methods suggesting their high reliability. Between the two new methods, the DL-FG method outperforms IN-FG, which was expected because DL-FG uses advanced deep learning methods for leakage correction, while IN-FG uses conventional SVM approach with a few hand-picked features.

C. Accuracy at Low CT Dosage

Accuracy results at low CT radiation dosage are presented in Figures 10 and 11 and Table 2. Figure 10 presents the reference mask and airway tree segmentation results for two subjects. For the subject at the top row, results by IS-BE show leakages near the upper-right and lower-left lobes which were manually removed, and branches were added by IS-AE near the lower-right lobe. Results using the two new methods are visually similar to corresponding reference masks with no apparent leakages. Also, both new methods captured additional branches near the lower- and upper-left lobes and lower-right lobe, which were missed by the industry-standard methods even after manual editing. For the case at the bottom row, there are no apparent leakages; but there are fewer branches as compared to the case at the top row. Among different methods, both IN-FG and DL-FG captured more peripheral airway branches than industry-standard methods, e.g., upper and lower left lobes.

Figure 11 presents quantitative accuracy results. Mean and variability of branch-counts in reference masks are shown in (a), while the branch-level accuracies are presented in (b). The two new methods captured all branches up to 6^{th} generation while IS-BE and IS-AE captured all branches only up to 5^{th} generation using low-dose CT input data. At 8^{th} generation, where the peak branch-count occurs in reference masks, IS-BE, IS-AE, IN-FG, and DL-FG captured 88.1, 92.2, 99.5, and 98.7% of true branches, respectively. Moreover, Figure 11(a) shows that mean airway branch count at standard and low radiation reference masks are similar up 7^{th} generation, and the difference become visible only at 8^{th} generation, and it grows beyond afterward. These results suggest that low radiation CT scans are suitable for airway analysis up to the segmental level and two generations beyond. However, the performance may not be optimum for peripheral airways.

For the 12 subjects in $Data_{low}$, the four methods successfully captured all five standard bronchi at the segmental level. IS-AE and the two new methods were able to capture all expected sub-segmental branches along the five standardized bronchial paths. Two sub-segmental branches along RB1, RB4, and LB1 were missed by IS-BE, which were manually added in IS-AE results. At the sub-sub-segmental level, both IN-FG and DL-FG missed only two branches, while the industry-standard methods missed 51 and 9 branches before and after editing. Like the results at standard dosage, the performance of IN-FG, DL-FG, and IS-AE along five standard bronchial paths are similar with small improvements using the new methods, especially at the sub-sub-segmental level.

Table 2 summarizes the results of leakage analysis. Prior to editing, on an average, the industry-standard method generated 10.2 leakages at a volume of 30 mm^3 per image, which were reduced to 0.4 leakages at 7 mm^3 per image after editing. Both IN-FG and DL-FG produced 0.3 or fewer leakage at low leakage volume per image. The largest leakage observed for DL-FG was 17 mm^3. In general, fewer and smaller leakages were observed in low radiation results as compared to those at standard radiation dosage. This observation is intriguing, and a possible justification is that leakages are often formed at higher generations with thinner airway walls, and, in low radiation scans, segmentation methods stop at relatively lower generations producing fewer leakages.

IV. Conclusions

CT-based fully automated algorithms for human airway tree segmentation have been presented and validated that at least match a method involving manual review and editing of an industry standard segmentation approach. [2] A multi-parametric FG segmentation algorithm has been developed, which provides a novel iterative framework that starts with a conservative parameter and progressively relaxes it until convergence. Effectiveness of the FG algorithm in iteratively augmenting airway branches using its parameter relaxation model has been demonstrated. Also, it has been experimentally observed that, in terms of branch-level accuracy and leakage performance, the fully automated intensity-based FG algorithm is superior to the manual edited results. A short coming of the FG algorithm is that it requires approximately one hour per CT scan.

To overcome the limitations of the CT intensity-based FG algorithm, while benefitting from the multi-parametric framework, a hybrid method combining DL-based airway lumen enhancement and FG algorithm has been developed. Specifically, a 3-D U-Net has been developed, trained, and validated for computing voxel-level likelihood maps of airway lumen from chest CT images. It has been shown that overlapping sub-regions at runtime reduces sub-region boundary artifacts and uncertainty. Like the intensity-based version, DL-based FG algorithm, also, has results superior to the manually edited results. DL-based FG algorithm has significantly improved the computational efficiency requiring approximately 6.5 minutes per CT scan. A major challenge with DL-based methods is related to their generalizability. Both training and testing data for our experiments were randomly chosen from ten different data acquisition sites with different CT scanners and imaging protocols. Also, the method has been applied to a dataset acquired with a different scan protocol at significantly reduced radiation. Although low radiation data were not used for training, experimental results were better than the industry-standard method with editing. These observations suggest that our DL-based FG method is generalizable.

Considering the performance and efficiency, the DL-based FG method is a viable fully automated solution for airway segmentation. Although the intensity-based FG algorithm is marginally superior to the DL-based version in terms of branch-level accuracy, its computation complexity makes it less suitable. An ideal setup for a large study will be to generate airway lumen masks for a smaller training dataset using the intensity-based FG algorithm; use the results to train a DL-based airway lumen classifier; and finally, apply a DL-based FG algorithm for larger study data.

Acknowledgements

This work was supported by the NIH grants R01 HL142042, R01 HL112986 and S10 OD018526. The authors thank the SPIROMICS participants and participating physicians, investigators and staff for making this research possible. More information about the study and how to access SPIROMICS data is at www.spiromics.org. We would like to acknowledge the following current and former investigators of the SPIROMICS sites and reading centers: Neil E Alexis, MD; Wayne H Anderson, PhD; Mehrdad Arjomandi, MD; Igor Barjaktarevic, MD, PhD; R Graham Barr, MD, DrPH; Lori A Bateman, MSc; Surya P Bhatt, MD; Eugene R Bleecker, MD; Richard C Boucher, MD; Russell P Bowler, MD, PhD; Stephanie A Christenson, MD; Alejandro P Comellas, MD; Christopher B Cooper, MD, PhD; David J Couper, PhD; Gerard J Criner, MD; Ronald G Crystal, MD; Jeffrey L Curtis, MD; Claire M Doerschuk, MD; Mark T Dransfield, MD; Brad Drummond, MD; Christine M Freeman, PhD; Craig Galban, PhD; MeiLan K Han, MD, MS; Nadia N Hansel, MD, MPH; Annette T Hastie, PhD; Eric A Hoffman, PhD; Yvonne Huang, MD; Robert J Kaner, MD; Richard E Kanner, MD; Eric C Kleerup, MD; Jerry A Krishnan, MD, PhD; Lisa M LaVange, PhD; Stephen C Lazarus, MD; Fernando J Martinez, MD, MS; Deborah A Meyers, PhD; Wendy C Moore, MD; John D Newell Jr, MD; Robert Paine, III, MD; Laura Paulin, MD, MHS; Stephen P Peters, MD, PhD; Cheryl Pirozzi, MD; Nirupama Putcha, MD, MHS; Elizabeth C Oelsner, MD, MPH; Wanda K O'Neal, PhD; Victor E Ortega, MD, PhD; Sanjeev Raman, MBBS, MD; Stephen I. Rennard, MD; Donald P Tashkin, MD; J Michael Wells, MD; Robert A Wise, MD; and Prescott G Woodruff, MD, MPH. The project officers from the Lung Division of the National Heart, Lung, and Blood Institute were Lisa Postow, PhD, and Lisa Viviano, BSN; SPIROMICS was supported by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, HHSN268200900020C), grants from the NIH/NHLBI (U01 HL137880 and U24 HL141762), and supplemented by contributions made through the Foundation for the NIH and the COPD Foundation from AstraZeneca/MedImmune; Bayer; Bellerophon Therapeutics; Boehringer-Ingelheim Pharmaceuticals, Inc.; Chiesi Farmaceutici S.p.A.; Forest Research Institute, Inc.; GlaxoSmithKline; Grifols Therapeutics, Inc.; Ikaria, Inc.; Novartis Pharmaceuticals Corporation; Nycomed GmbH; ProterixBio; Regeneron Pharmaceuticals, Inc.; Sanofi; Sunovion; Takeda Pharmaceutical Company; and Theravance Biopharma and Mylan. The authors wish to thank Melissa Saylor, Jarron Atha, Patrick Thalken, and Mark Esher for research assistance and data transfer, and Kálmán Palágyi for providing his skeletonization algorithm.

This work was supported by the NIH grants R01-HL112986 and S10-OD026960.

References

[1]. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al., "Genetic epidemiology of COPD (COPDGene) study design," COPD, vol. 7, no. 1, pp. 32–43, 2011.

[2]. Sieren JP, Newell JD Jr., Barr RG, Bleecker ER, Burnette N, Carretta EE, et al., "SPIROMICS Protocol for Multicenter Quantitative Computed Tomography to Phenotype the Lungs," Am J Respir Crit Care Med, vol. 194, no. 7, pp. 794–806, 10 1, 2016. [PubMed: 27482984]

[3]. Couper D, LaVange LM, Han M, Barr RG, Bleecker E, Hoffman EA, et al., "Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS)," Thorax, vol. 69, no. 5, pp. 491–4, 5 2014.

[4]. Donohue KM, Hoffman EA, Baumhauer H, Guo J, Budoff M, Austin JH, et al., "Cigarette smoking and airway wall thickness on CT scan in a multi-ethnic cohort: the MESA Lung Study," Respir Med, vol. 106, no. 12, pp. 1655–64, 12 2012. [PubMed: 22974831]

[5]. Bourbeau J, Tan WC, Benedetti A, Aaron SD, Chapman KR, Coxson HO, et al., "Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the need for longitudinal observational studies in COPD," COPD, vol. 11, no. 2, pp. 125–32, 4, 2014. [PubMed: 22433011]

[6]. Castro M, Fain SB, Hoffman EA, Gierada DS, Erzurum SC, Wenzel S, et al., "Lung imaging in asthmatic patients: the picture is clearer," J Aller Clin Immun, vol. 128, no. 3, pp. 467–478, 2011.

[7]. Smith BM, Hoffman EA, Rabinowitz D, Bleecker E, Christenson S, Couper D, et al., "Comparison of spatially matched airways reveals thinner airway walls in COPD. The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study and the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS)," Thorax, vol. 69, no. 11, pp. 987–96, 11, 2014. [PubMed: 24928812]

[8]. Smith BM, Traboulsi H, Austin JHM, Manichaikul A, Hoffman EA, Bleecker ER, et al., "Human airway branch variation and chronic obstructive pulmonary disease," Proc Natl Acad Sci U S A, vol. 115, no. 5, pp. E974–E981, 1 30, 2018. [PubMed: 29339516]

[9]. Haghighi B, Choi S, Choi J, Hoffman EA, Comellas AP, Newell JD Jr., et al., "Imaging-based clusters in current smokers of the COPD cohort associate with clinical characteristics: the SubPopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS)," Respir Res, vol. 19, no. 1, pp. 178, 9 18, 2018. [PubMed: 30227877]

[10]. Kirby M, Tanabe N, Tan WC, Zhou G, Obeidat M, Hague CJ, et al., "Total Airway Count on Computed Tomography and the Risk of Chronic Obstructive Pulmonary Disease Progression. Findings from a Population-based Study," Am J Respir Crit Care Med, vol. 197, no. 1, pp. 56–65, 1 1, 2018. [PubMed: 28886252]

[11]. Tschirren J, Hoffman EA, McLennan G, and Sonka M, "Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans," IEEE Trans Med Imag, vol. 24, no. 12, pp. 1529–39, 12 2005.

[12]. Tschirren J, McLennan G, Palagyi K, Hoffman EA, and Sonka M, "Matching and anatomical labeling of human airway tree," IEEE Trans Med Imag, vol. 24, no. 11, pp. 1540–7, 12 2005.

[13]. Cootes TF, Taylor CJ, Cooper DH, and Graham J, "Active shape models-their training and application," Comput Vis Imag Und, vol. 61, no. 1, pp. 38–59, 1995.

[14]. Kass M, Witkin A, and Terzopoulos D, "Snakes: Active contour models," Int J Comput Vis, vol. 1, no. 4, pp. 321–331, 1988.

[15]. Udupa JK and Saha PK, "Fuzzy connectedness and image segmentation," Proc IEEE, vol. 91, no. 10, pp. 1649–1669, 2003.

[16]. Saha PK, Gao Z, Alford SK, Sonka M, and Hoffman EA, "Topomorphologic separation of fused isointensity objects via multiscale opening: separating arteries and veins in 3-D pulmonary CT," IEEE Trans Med Imag, vol. 29, no. 3, pp. 840–51, 3 2010.

[17]. Aysola RS, Hoffman EA, Gierada D, Wenzel S, Cook-Granroth J, Tarsi J, et al., "Airway remodeling measured by multidetector CT is increased in severe asthma and correlates with pathology," Chest, vol. 134, no. 6, pp. 1183–1191, 2008. [PubMed: 18641116]

[18]. Lo P, van Ginneken B, Reinhardt JM, Yavarna T, de Jong PA, Irving B, et al., "Extraction of airways from CT (EXACT'09)," IEEE Trans Med Imag, vol. 31, no. 11, pp. 2093–107, 11 2012.

[19]. van Rikxoort EM and van Ginneken B, "Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review," Phys Med Biol, vol. 58, no. 17, pp. R187–220, 9 7 2013. [PubMed: 23956328]

[20]. Pu J, Gu S, Liu S, Zhu S, Wilson D, Siegfried JM, et al., "CT based computerized identification and analysis of human airways: a review," Med Phys, vol. 39, no. 5, pp. 2603–16, 5 2012. [PubMed: 22559631]

[21]. Schlathoelter T, Lorenz C, Carlsen IC, Renisch S, and Deschamps T, "Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy," in Proc SPIE Med Imag: Imag Proc, 2002, pp. 103–114.

[22]. Kiraly AP, Higgins WE, McLennan G, Hoffman EA, and Reinhardt JM, "Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy," Acad Radiol, vol. 9, no. 10, pp. 1153–68, 10 2002. [PubMed: 12385510]

[23]. Fabijańska A, "Two-pass region growing algorithm for segmenting airway tree from MDCT chest scans," Comput Med Imag Graph, vol. 33, no. 7, pp. 537–546, 2009.

[24]. Graham MW, Gibbs JD, Cornish DC, and Higgins WE, "Robust 3-D airway tree segmentation for image-guided peripheral bronchoscopy," IEEE Trans Med Imag, vol. 29, no. 4, pp. 982–97, 4 2010.

[25]. Aykac D, Hoffman EA, McLennan G, and Reinhardt JM, "Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images," IEEE Trans Med Imag, vol. 22, no. 8, pp. 940–50, 8 2003.

[26]. Fetita C, Ortner M, Brillet P-Y, Prêteux F, and Grenier P, "A morphological-aggregative approach for 3D segmentation of pulmonary airways from generic MSCT acquisitions," in Proc 2nd Int Workshop Pulmo Imag Anal, 2009, pp. 215–226.

[27]. Sonka M, Park W, and Hoffman EA, "Rule-based detection of intrathoracic airway trees," IEEE Trans Med Imag, vol. 15, no. 3, pp. 314–26, 1996.

[28]. Kitasaka T, Mori K, Hasegawa J.-i., Suenaga Y, and Toriwaki J.-i., “Extraction of bronchus regions from 3D chest X-ray CT images by using structural features of bronchus,” in Proc Int Cong Comput Assist Rad Surg (CARS), 2003, pp. 240–245.

[29]. Meng Q, Kitasaka T, Nimura Y, Oda M, Ueno J, and Mori K, “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, vol. 12, no. 2, pp. 245–261, 2 2017. [PubMed: 27796791]

[30]. van Ginneken B, Baggerman W, and van Rikxoort EM, “Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans,” in Proc Med Imag Comp Comp-Assist Interv (MICCAI), 2008, pp. 219–226.

[31]. Deng L and Yu D, “Deep learning: methods and applications,” Found Trend Sig Proc, vol. 7, no. 3–4, pp. 197–387, 2014.

[32]. Goodfellow I, Bengio Y, and Courville A, Deep learning: MIT Press, 2016.

[33]. LeCun Y, Bengio Y, and Hinton G, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–44, 5 28 2015. [PubMed: 26017442]

[34]. Charbonnier J-P, Van Rikxoort EM, Setio AA, Schaefer-Prokop CM, van Ginneken B, and Ciompi F, “Improving airway segmentation in computed tomography using leak detection with convolutional networks, ” Med Imag Anal, vol. 36, pp. 52–60, 2017.

[35]. Krizhevsky A, Sutskever I, and Hinton GE, “Imagenet classification with deep convolutional neural networks, ” in Proc Adv Neural Inf Proc Sys, 2012, pp. 1097–1105.

[36]. Jin D, Xu Z, Harrison AP, George K, and Mollura DJ, “3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels, ” in Proc Int Workshop Mach Learn Med Imag, 2017, pp. 141–149.

[37]. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, and Ronneberger O, “3D U-Net: learning dense volumetric segmentation from sparse annotation, ” in Proc Med Imag Comp Comp-Assist Interv (MICCAI), 2016, pp. 424–432.

[38]. Ronneberger O, Fischer P, and Brox T, “U-net: Convolutional networks for biomedical image segmentation,” in Proc Med Imag Comp Comp-Assist Interv (MICCAI), 2015, pp. 234–241.

[39]. Yun J, Park J, Yu D, Yi J, Lee M, Park HJ, et al., “Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net,” Med Imag Anal, vol. 51, pp. 13–20, 2019.

[40]. Juarez AG-U, Tiddens HA, and de Bruijne M, “Automatic airway segmentation in chest CT using convolutional neural networks, ” in Proc 3rd Int Workshop Imag Anal Mov Organ, Breast, Thoracic Imag, ed: Springer, 2018, pp. 238–250.

[41]. Nadeem SA, Hoffman EA, Sieren JP, and Saha PK, “Topological leakage detection and freeze-and-grow propagation for improved CT-based airway segmentation,” in Proc SPIE Med Imag: Imag Proc, 2018, p. 105741A.

[42]. Cortes C and Vapnik V, “Support-vector networks,” Mach Learn, vol. 20, no. 3, pp. 273–297, 1995.

[43]. Mori K, Hasegawa J.-i., Toriwaki J.-i., Anno H, and Katada K, “Recognition of bronchus in three-dimensional X-ray CT images with applications to virtualized bronchoscopy system,” in Proc 13th Int Conf Pat Rec (ICPR), 1996, pp. 528–532.

[44]. Saha PK, Wehrli FW, and Gomberg BR, “Fuzzy distance transform: theory, algorithms, and applications,” Comput Vis Imag Und, vol. 86, no. 3, pp. 171–190, 2002.

[45]. Nadeem SA, Hoffman EA, and Saha PK, “An automated airway segmentation algorithm for CT images using topological leakage detection and volume freezing,” in Proc 24th Int Conf Pat Rec (ICPR), 2018, pp. 1181–1186.

[46]. Palágyi K, “A Sequential 3D Curve-Thinning Algorithm Based on Isthmuses,” in Proc Int Symp Vis Comput (ISVC), 2014, pp. 406–415.

[47]. Fuld MK, Grout RW, Guo J, Morgan JH, and Hoffman EA, “Systems for lung volume standardization during static and dynamic MDCT-based quantitative assessment of pulmonary structure and function,” Acad Radiol, vol. 19, no. 8, pp. 930–40, 8, 2012. [PubMed: 22555001]

[48]. Iyer KS, Grout RW, Zamba GK, and Hoffman EA, “Repeatability and Sample Size Assessment Associated with Computed Tomography-Based Lung Density Metrics,” COPD, vol. 1, no. 1, pp. 97–104, 2014. [PubMed: 25553338]

[49]. Kingma DP, and Ba J, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

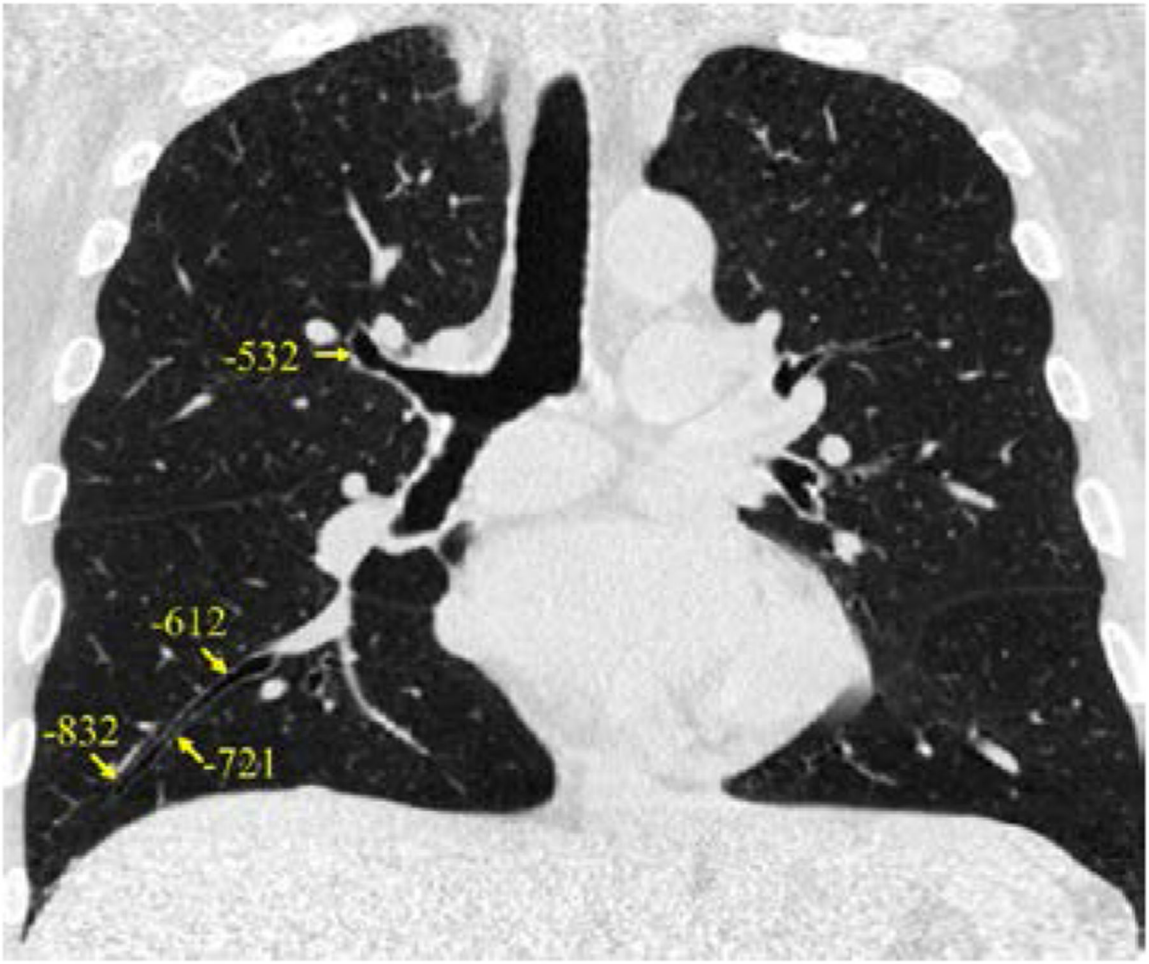

Figure 1.

Decaying CT intensities of airway walls from proximal to distal sites in a TLC CT scan. At proximal airway sites, high CT intensity of −456 HU or similar values are observed for airway walls, which are reduced to as low as −832 HU at distal sites.

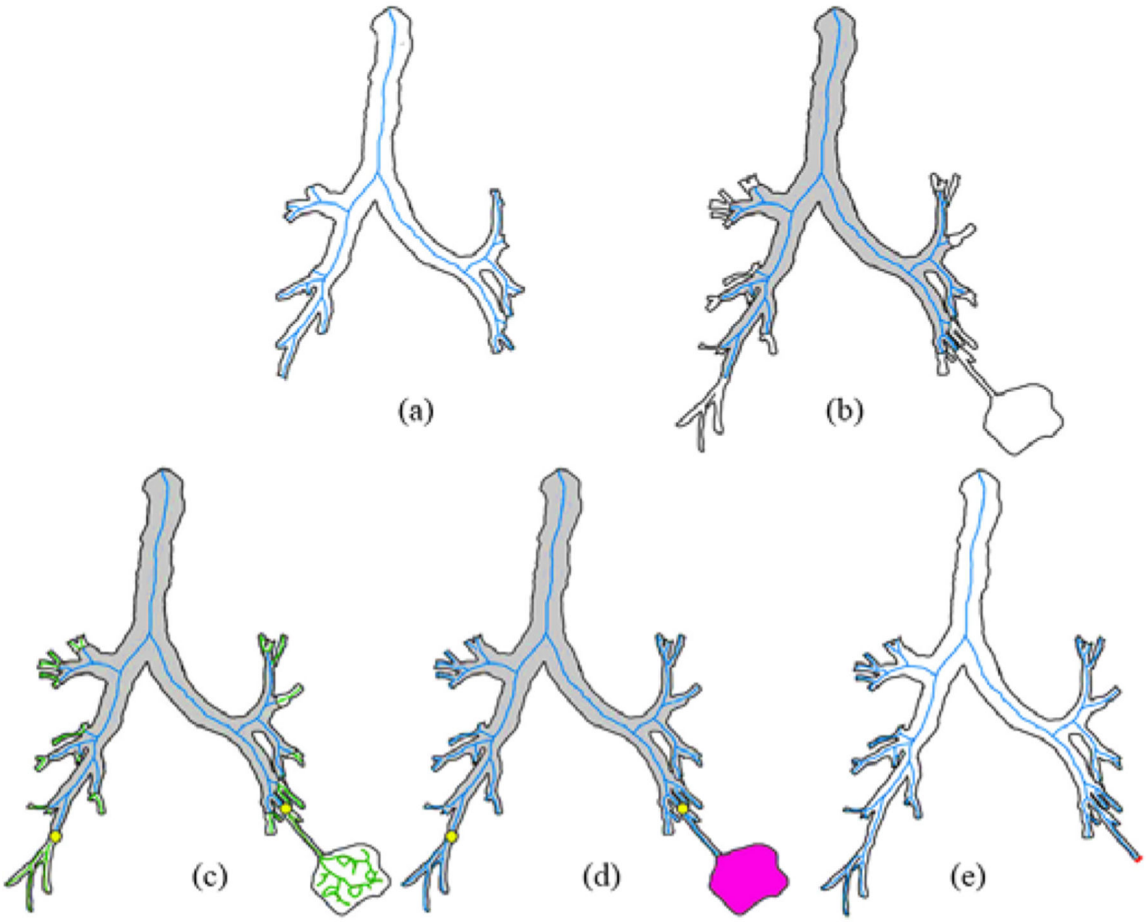

Figure 2.

Illustration of intermediate steps during an iteration of the FG algorithm. (a) Confident airway volume CAV and its centerline (blue) at the beginning of an iteration. (b) Initial region growing from CAV. (c) Identification of possible leakages (yellow dots) by locating regions with rapid growth (centerline in green). (d) Confirmation of leakages (magenta). (e) Augmentation of leakage-free CAV and insertion of forbidden volume (red) around the leakage-root. The algorithm enters the next iteration with the result of (e) as input i.e. CAV, its skeleton, and FV.

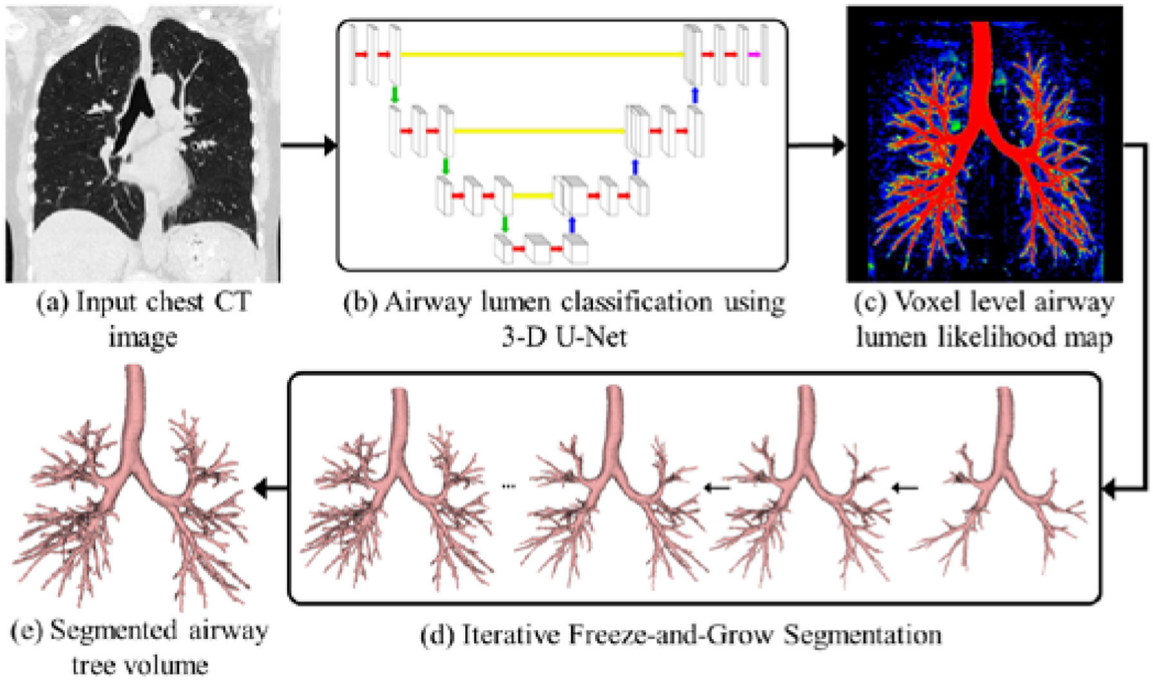

Figure 3.
Block diagram of the major steps of the deep learning-based FG airway segmentation algorithm.

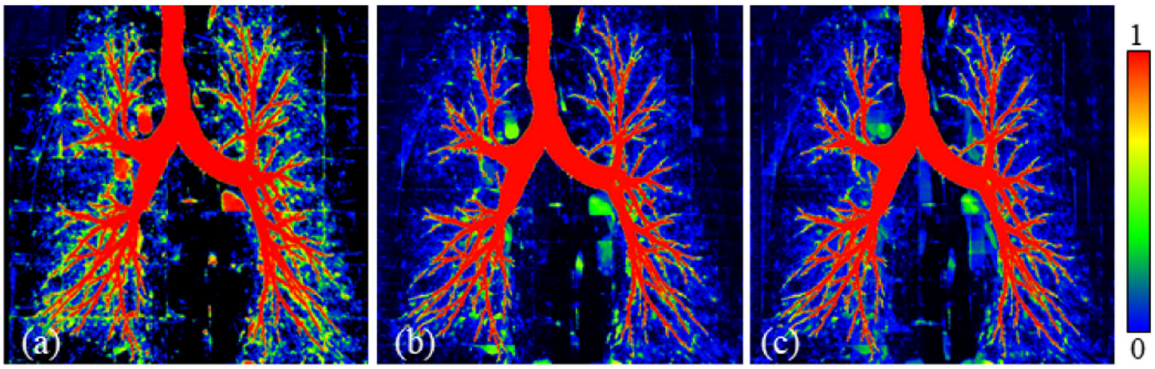

Figure 4.

Results of DL-based voxel-level airway lumen likelihood classification. (a-c) Color-coded MIPs of voxel-level likelihood maps generated using our modified 3-D U-Net. Sample grid spacings of 64, 32, and 16 voxels were used for (a-c), respectively. Checkerboard-type sub-region boundary artifacts are visible in (a), which mostly disappear in (b,c).

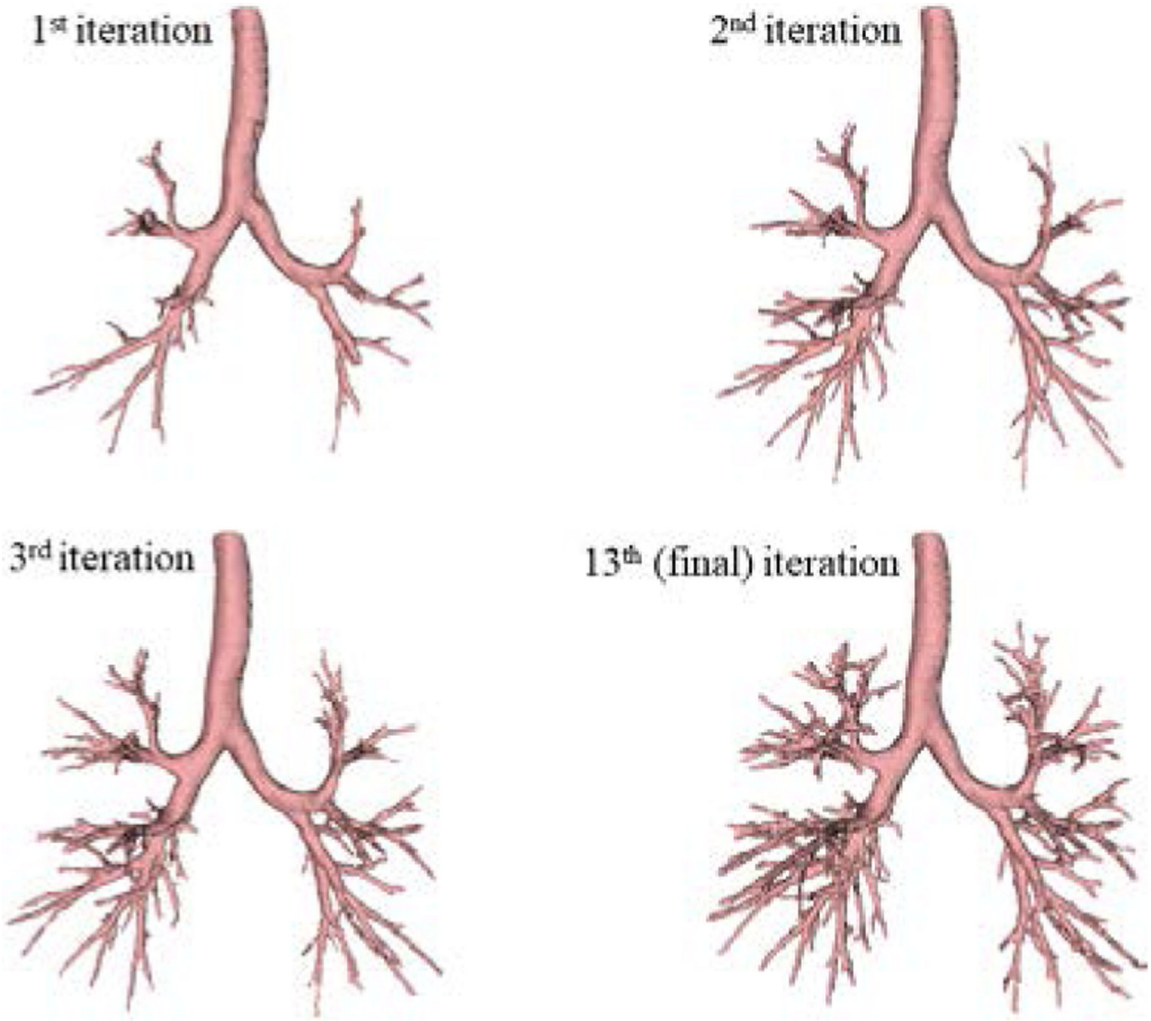

Figure 5.
Intermediate segmentation results of airway lumen segmentation at different FG iterations using the DL-FG algorithm.

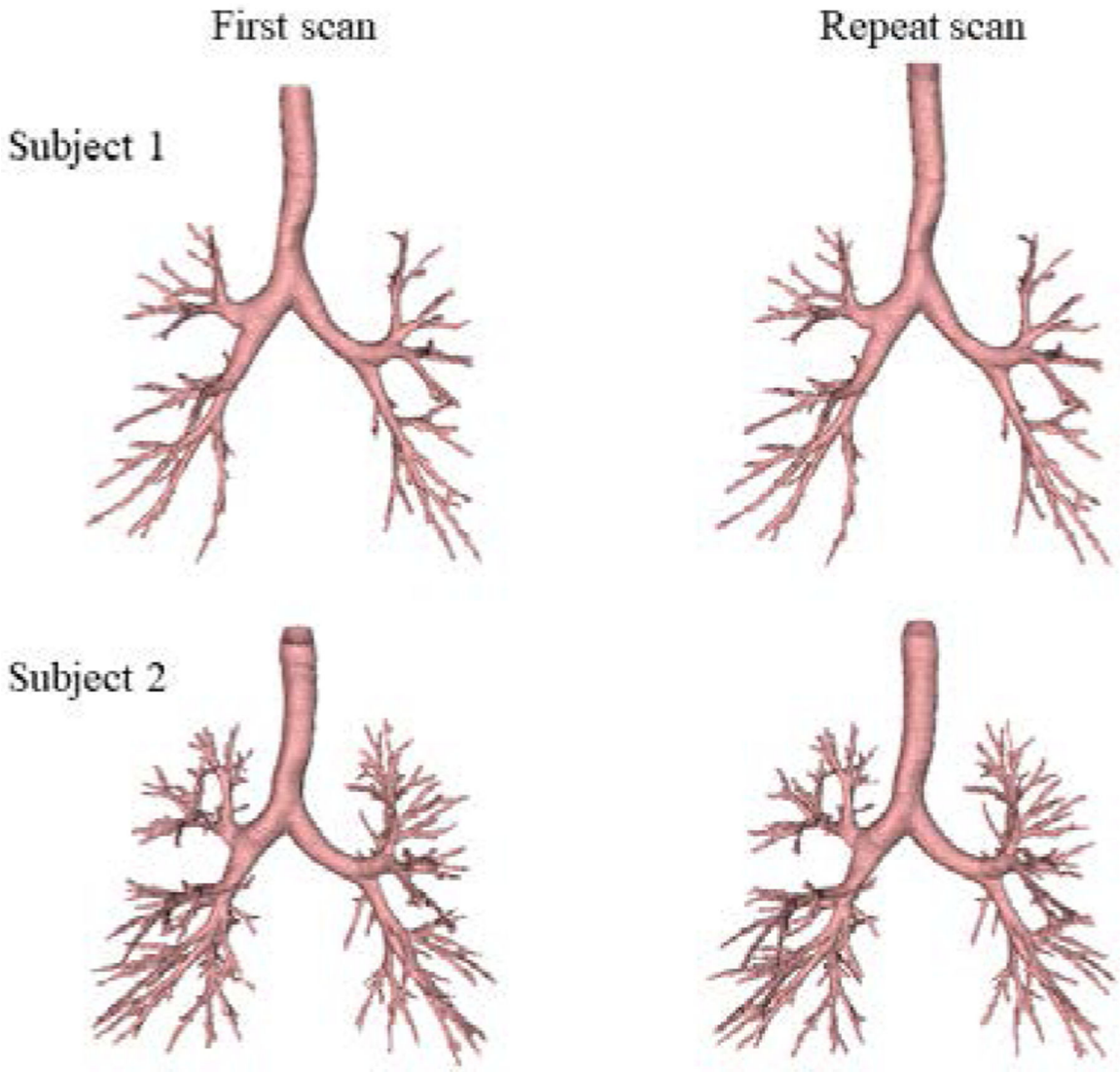

Figure 6.
Repeat scan reproducibility of CT-based airway lumen segmentation for two different human subjects using the DL-FG algorithm.

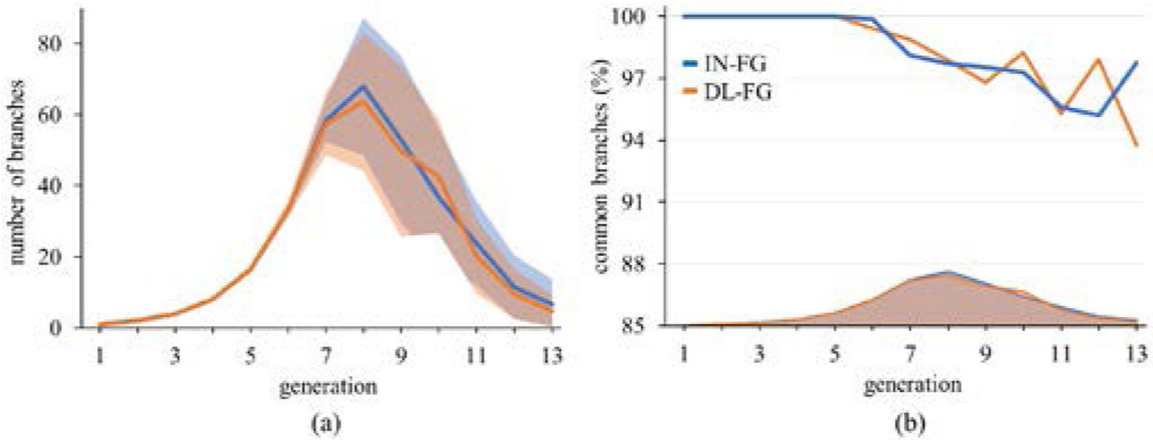

Figure 7.

Quantitative repeat CT scan reproducibility oi airway segmentation using IN-FG and DL-FG. (a) Mean (solid line) and variance (shaded region) of segmented airway counts at different generations. (b) Number of common segmented branches in repeat scans at different generations. Mean branch-counts of (a) are copied at the bottom.

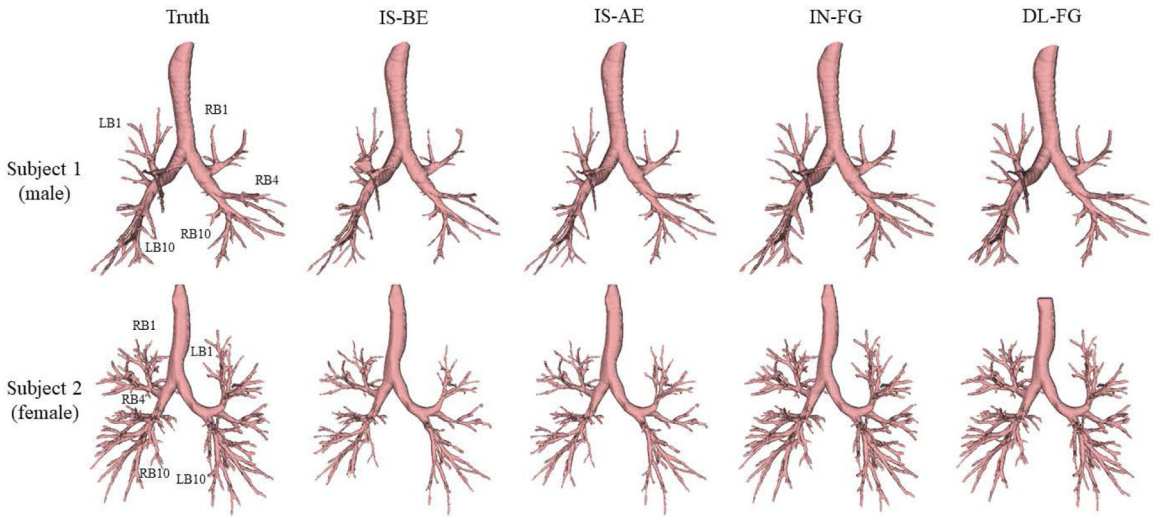

Figure 8.

Accuracy of different airway segmentation methods as compared to the reference mask. Results from two subjects are shown on two rows.

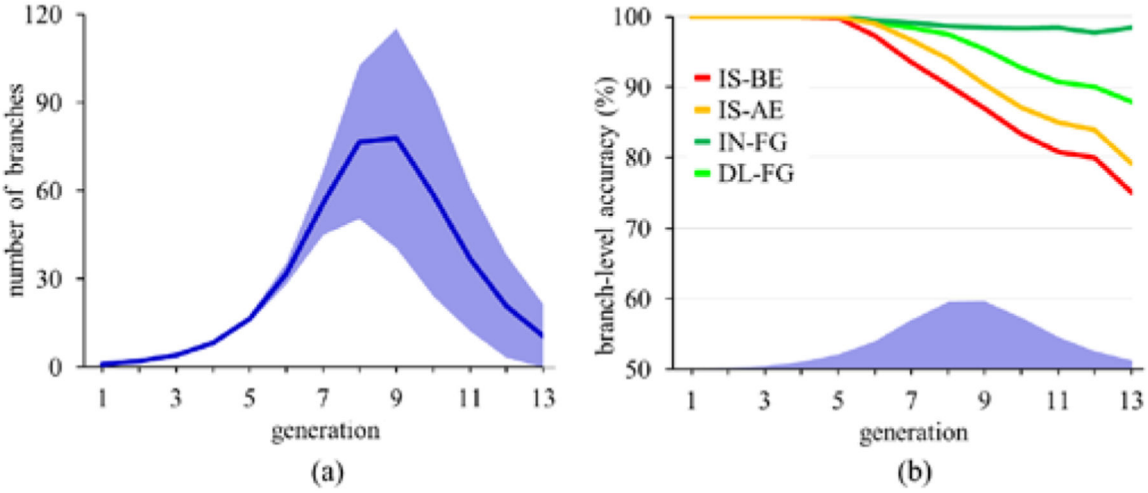

Figure 9.

Quantitative accuracy of airway tree segmentation at standard radiation CT scans. (a) Mean and variance of airway counts in reference segmentation masks at different generations. (b) Percentage of true branches captured by different methods. Mean branch-counts of (a) are copied at the bottom.

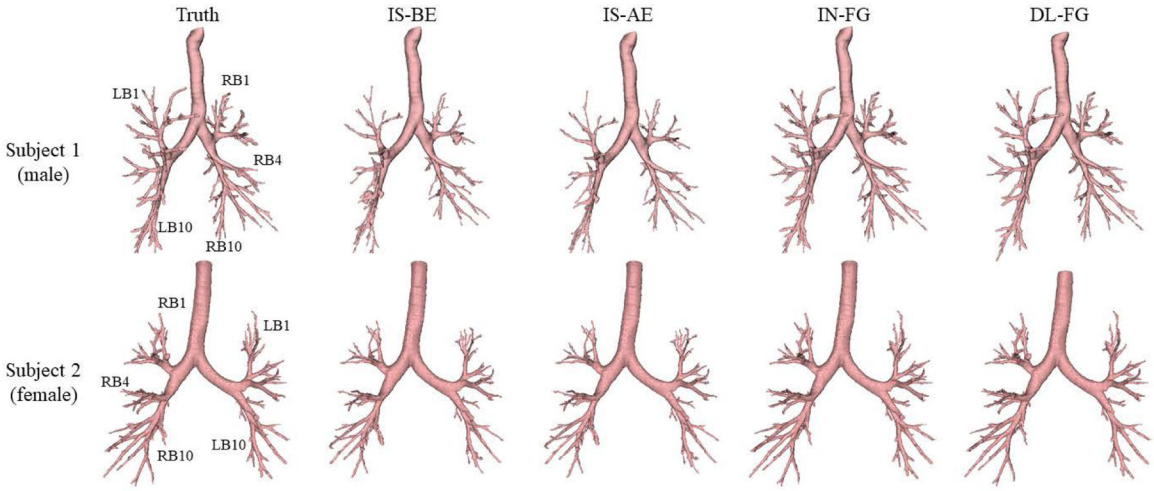

Figure 10.

Accuracy of different airway segmentation methods at low CT radiation. Results from two subjects are shown on two rows.

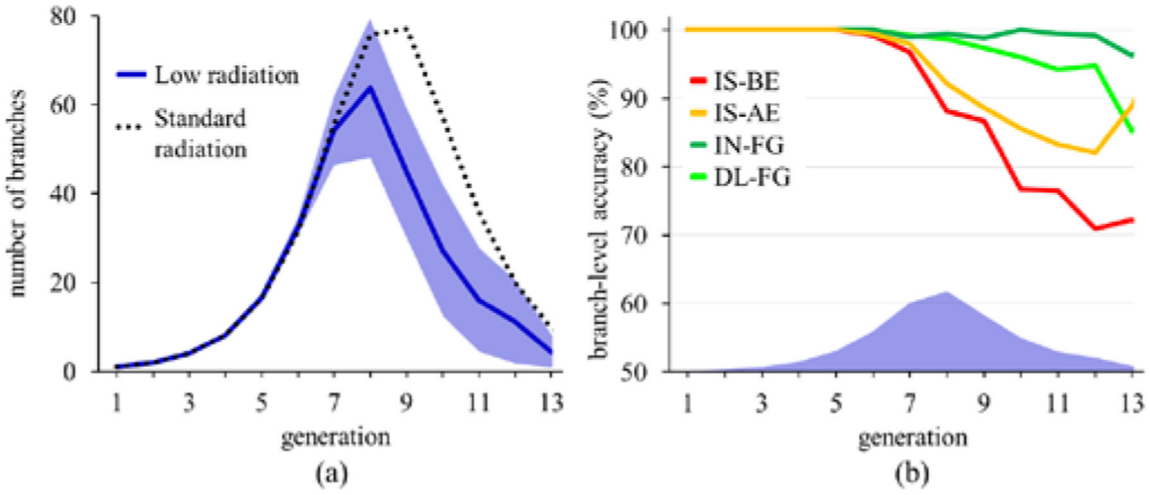

Figure 11.

Quantitative accuracy of airway tree segmentation for low radiation CT scans. (a) Mean (blue line) and variance (shaded region) of airway counts at different generations in reference segmentation masks for low radiation CT scans. Mean airway branch counts of reference masks at standard CT radiation are shown by the dotted line. (b) Percentage of true branches captured by different methods.

TABLE I

Statistics of leakages by different airway segmentation methods

	IS-BE	IS-AE	IN-FG	DL-FG
Leakage count	7.4±8.4	1.2±1.6	0.7±1.0	0.4±0.8
Total leakage volume (mm^3) (mean±std.)	449±1143	54±185	20±40	10±23
Largest leakage size (mm^3)	7571	1642	173	29

TABLE II

Statistics of leakages by different airway segmentation methods at low radiation CT scans

	IS-BE	IS-AE	IN-FG	DL-FG
Leakage Count	10.2±4.0	0.4±0.5	0.3±0.5	0.1±0.0
Total Leakage Volume (mm^3) (mean±std.)	30±25	7±9	6±9	l±0.0
Largest leakage size (mm^3)	98	21	19	17