

# Airway Detection in COPD at Low-Dose CT Using Deep Learning and Multiparametric Freeze and Grow

Syed Ahmed Nadeem, PhD • Alejandro P. Comellas, MD • Eric A. Hoffman, PhD • Punam K. Saba, PhD

From the Departments of Radiology (S.A.N., E.A.H., P.K.S.) and Internal Medicine (A.P.C., E.A.H.), Carver College of Medicine, and Departments of Biomedical Engineering (E.A.H.) and Electrical and Computer Engineering (P.K.S.), College of Engineering, The University of Iowa, 4312 Seamans Center, Iowa City, IA 52242. Received December 12, 2021; revision requested February 15, 2022; revision received September 27; accepted October 27. Address correspondence to P.K.S. (email: punam-saba@uiowa.edu).

Supported by the National Institutes of Health (grants R01 HL142042, 5U01 HL089897, and R01 HL112986).

Conflicts of interest are listed at the end of this article.

Radiology: Cardiothoracic Imaging 2022; 4(6):e210311 • <https://doi.org/10.1148/ryct.210311> • Content codes: **CH** **CT**

**Purpose:** To present and validate a fully automated airway detection method at low-dose CT in patients with chronic obstructive pulmonary disease (COPD).

**Materials and Methods:** In this retrospective study, deep learning (DL) and freeze-and-grow (FG) methods were optimized and applied to automatically detect airways at low-dose CT. Four data sets were used: two data sets consisting of matching standard- and low-dose CT scans from the Genetic Epidemiology of COPD (COPDGene) phase II (2014–2017) cohort ( $n = 2 \times 236$ ; mean age  $\pm$  SD, 70 years  $\pm$  9; 123 women); one data set consisting of low-dose CT scans from the COPDGene phase III (2018–2020) cohort ( $n = 335$ ; mean age  $\pm$  SD, 73 years  $\pm$  8; 173 women); and one data set consisting of low-dose, anonymized CT scans from the 2003 Dutch–Belgian Randomized Lung Cancer Screening trial ( $n = 55$ ) acquired by using different CT scanners. Performance measures for different methods were computed and compared by using the Wilcoxon signed rank test.

**Results:** At low-dose CT, 56 294 of 62 480 (90.1%) airways of the reference total airway count (TAC) and 32 109 of 37 864 (84.8%) airways of the peripheral TAC (TAC<sub>p</sub>), detected at standard-dose CT, were detected. Significant losses ( $P < .001$ ) of 14 526 of 76 453 (19.0%) airways and 884 of 6908 (12.8%) airways in the TAC and 12 256 of 43 462 (28.2%) airways and 699 of 3882 (18.0%) airways in the TAC<sub>p</sub> were observed, respectively, for the multiprotocol and multiscanner data without retraining. When using the automated low-dose CT method, TAC values of 347, 342, 323, and 266 and TAC<sub>p</sub> values of 205, 202, 289, and 141 were observed for those who have never smoked and participants at Global Initiative for Chronic Obstructive Lung Disease stages 0, 1, and 2, respectively, which were superior to the respective values previously reported for matching groups when using a semiautomated method at standard-dose CT.

**Conclusion:** A low-cost, automated CT-based airway detection method was suitable for investigation of airway phenotypes at low-dose CT.

Clinical trial registration no. NCT00608764

Supplemental material is available for this article.

© RSNA, 2022

Chronic obstructive pulmonary disease (COPD) is a common inflammatory disease characterized by airflow limitation and is one of the leading causes of death in the United States (1–3). The current paradigm suggests that early signs of COPD originate with changes in airway structure and physiologic function (4,5); specifically, small airways are the first to be affected with narrowing and loss (6,7). Moreover, studies have shown that common airway branching variants confer a greater risk of acquiring COPD in both smoking and nonsmoking populations (8). Therefore, airway phenotypes have become important indicators for understanding the mechanisms of disease, severity, and progression in quantitative CT-based studies (9–13). Airway segmentation is a precursor for CT-based analysis of airway phenotypes, but state-of-the-art methods require manual review and correction for missing branches and segmentation leakages (10,14), which are a bottleneck for multicenter longitudinal lung studies. Moreover, using low-dose CT imaging is the current trend in most

CT-based studies because it reduces participants' cumulative exposure to ionizing radiation and its associated risks (15). For example, only low-dose CT scans are acquired at phase III visits of the Genetic Epidemiology of COPD (COPD-Gene) study (phase III:  $n = 2243$  as of September 2020). The increased noise and reduced image quality of low-dose CT scans (16) reduces the accuracy in measurement of CT-based lung phenotypes (17) and adds further challenges for detecting airways, especially small airways (18). Thus, an automated airway detection method at low-dose CT, which would provide results comparable or superior to those previously reported at standard-dose CT, could be an immensely powerful tool for CT-based lung studies.

This article presents a low-dose CT-based airway detection algorithm that uses deep learning (DL) and freeze-and-grow (FG) methods that are suitable for large studies acquiring low-dose chest CT scans. The FG method was previously validated for standard-dose chest CT scans, and details of the method were presented (19). The basic

## Abbreviations

COPD = chronic obstructive pulmonary disease, COPDGene = Genetic Epidemiology of COPD, Data  $C_{II-Low}$  = low-dose CT scans from phase II visits of COPDGene Iowa cohort, Data  $C_{II-Stand}$  = standard-dose CT scans from phase II visits of the COPDGene Iowa cohort, Data  $C_{III-Low}$  = low-dose CT scans from phase III visits of the COPDGene Iowa cohort, Data  $N_{Low}$  = low-dose chest CT scans from the Dutch–Belgian Randomized Lung Cancer Screening trial, DL = deep learning, FG = freeze and grow, GOLD = Global Initiative for Chronic Obstructive Lung Disease, TAC = total airway count,  $TAC_p$  = peripheral TAC, TLC = total lung capacity

## Summary

When using the fully automated deep learning and multiparametric freeze-and-grow method, low-dose CT depicted most total and peripheral airways found at standard-dose CT in patients with chronic obstructive pulmonary disease.

## Key Points

- At low-dose CT, 56 294 of 62 480 (90.1%) of total and 32 109 of 37 864 (84.8%) of peripheral reference airways, detected at standard-dose CT, were successfully detected.
- Improved performance ( $P < .001$ ) of 12 256 of 43 462 (28.2%) and 699 of 3882 (18.0%) peripheral airways was observed by retraining the method on multiprotocol and multiscanner low dose CT compared with a generalized method.
- The cost of 7 days of computer time and 10 hours of expert time for retraining the algorithm for multiprotocol and multiscanner data was low compared with performance loss ( $P < .001$ ).

## Keywords

Airway, Airway Count, Airway Detection, Chronic Obstructive Pulmonary Disease, CT, Deep Learning, Generalizability, Low-Dose CT, Segmentation, Thorax, Lung

principle of the method is to freeze leakages at the current threshold parameter and move to the next generous parameter to grow further. DL methods offer data-driven paradigms for deriving optimum multilayered and multiscale features for a target application without requiring ad hoc rules or process-level design, and they outperform conventional methods (19). However, fully data-driven approaches have inevitable limitations related to their generalizability to different data sets from similar applications. Experiments were designed to investigate the following: optimization of the DL-FG algorithm, the accuracy of DL-FG airway detection at low-dose CT, and the trade-off between the costs of retraining and the detection performance regarding the generalizability of multiprotocol and multiscanner data.

## Materials and Methods

### Study Design

This retrospective study was designed to optimize a DL-FG airway detection algorithm and examine its accuracy and multiprotocol and multiscanner generalizability at low-dose CT. Experiments were performed on total lung capacity (TLC) CT scans. Standard- and low-dose chest CT scans from phase II (2014–2017) and III (2018–2020) visits of the COPDGene (NCT00608764) Iowa cohort and low-dose CT scans from the Dutch–Belgian Randomized Lung Cancer Screening trial

(2003, NTR636) (20), the largest CT lung cancer screening trial in Europe, were used. The COPDGene Iowa cohort was approved by the University of Iowa Institutional Review Board, and written informed consent was obtained from all participants. The current study was Health Insurance Portability and Accountability Act compliant. COPDGene data include data from participants studied across multiple institutions, and a broad set of publications has been derived from the data gathered.

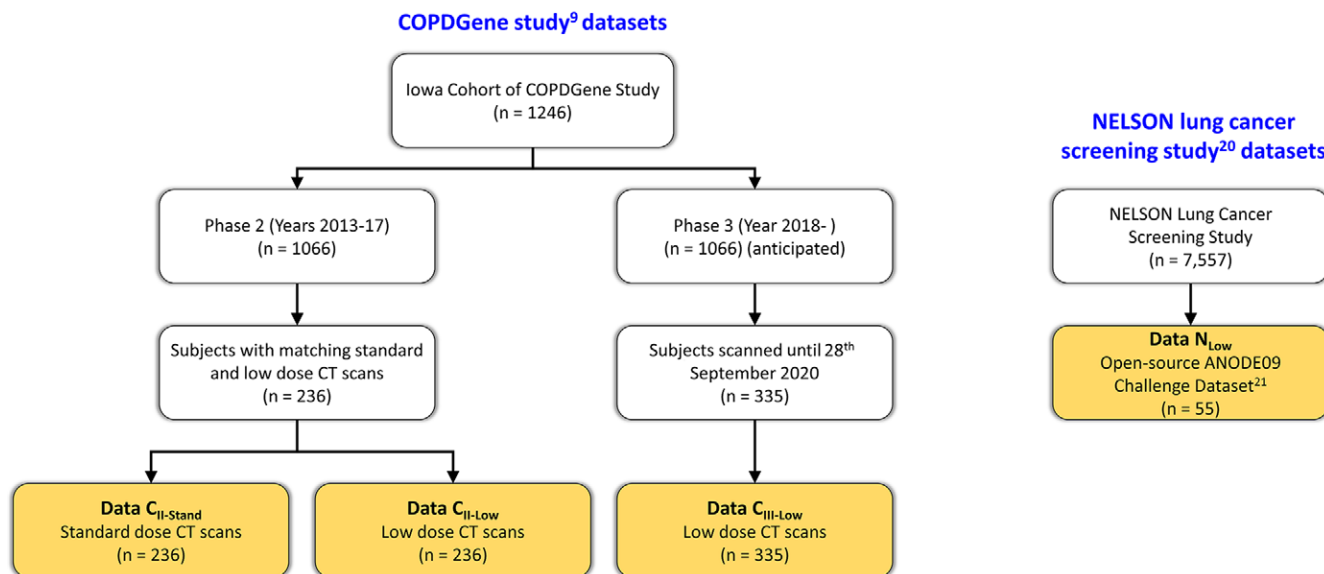
### Chest CT Scans

Figure 1 illustrates the selection flowcharts for the four chest CT data sets for the experiments. Standard- and low-dose CT data sets, Data  $C_{II-Stand}$  and  $C_{II-Low}$ , consisted of matching standard- and low-dose TLC CT scans of 236 participants (70 years  $\pm$  9 [SD]; 123 women) acquired at COPDGene phase II visits. The data set Data  $C_{III-Low}$  was prepared using low-dose CT scans of 335 participants (73 years  $\pm$  8; 173 women) acquired at COPDGene phase III visits. All CT scans were acquired with a Siemens SOMATOM Force scanner. Standard-dose CT scans were acquired at a volume CT dose index of 13.3 mGy, whereas low-dose scans were acquired at 2.2 mGy. These images were reconstructed on  $512 \times 512$  matrices at 0.5-mm slice spacing. Details of the CT imaging protocol were previously published (9,10). Data  $N_{Low}$  was collected from the Automatic Nodule Detection 2009 data set (21), which consisted of 55 anonymized CT scans that are publicly available online (<https://anode09.grand-challenge.org/>) and are a subset from the Dutch–Belgian Randomized Lung Cancer Screening trial. The scans were acquired with either a 16-slice or a 64-slice CT scanner (Philips Medical Systems) at 2.2 mGy and were reconstructed on  $512 \times 512$  matrices at 0.7-mm slice spacing; protocol details are available in literature (15).

### Automated Airway Detection Using DL and FG

Automated airway detection was completed in two major steps: airway tree segmentation and branch counting at different generations. A recent airway segmentation algorithm that uses DL and multiparametric FG methods (19) was optimized. The initial DL module was used to generate an airway lumen likelihood map from a chest CT image, which was fed into the iterative FG module, generating multiparametric segmentation of the airway tree volume. The DL module was implemented by using a three-dimensional U-Net with three pooling layers and three deconvolutional layers (22,23). The network was trained to output a voxel-level lumen likelihood map on a  $64 \times 64 \times 64$ -voxel CT input patch. The iterative FG module uses detection of leakages, forbidden volume marking and freezing around leakage roots by prohibiting connectivity paths, and parameter relaxation to facilitate further airway tree growth (19).

After airway segmentation, the tree volume was skeletonized by using a minimum cost path-based curve skeletonization algorithm producing a single-voxel thin representation (24,25). Spurious skeletal branches were automatically pruned by using local scale and tree depth at skeletal tree junctions (26). Finally,



**Figure 1:** Participant selection flowchart for each data set. ANODE09 = Automatic Nodule Detection 2009, COPDGene = Genetic Epidemiology of Chronic Obstructive Pulmonary Disease, NELSON = Dutch–Belgian Randomized Lung Cancer Screening.

airway counts at each generation were computed by using the trachea as generation 0.

### Experiments

Experiments were conducted to optimize the network and training data size for the DL-FG algorithm and evaluate algorithm accuracy and generalizability at low-dose CT. For different evaluative experiments, training, reference, and evaluation data sets were carefully tailored to accomplish the objective of each experiment, as illustrated in Figure 2. Training and evaluation data sets were mutually exclusive, and the reference segmentation of airway volume was generated by using CT intensity–based FG airway segmentation (19) and manual editing on an ITK-SNAP graphical user interface by a trained expert (S.A.N.) with 5 years of experience working on pulmonary CT imaging (27).

**DL-FG algorithm optimization.**— Network width (ie, the number of kernels at the first pooling layer) and training data size are key parameters defining the computational complexity of a DL network. The trade-offs between the performance and computational complexity of the DL module were assessed by using various combinations of network parameters to select an effective setup. The network width and training data size were varied between 40 and 80 kernels and 20 and 60 scans, respectively. A fixed evaluation data set of 100 scans from Data  $C_{II-Stand}$ , separate from all training data sets, was used for performance evaluation.

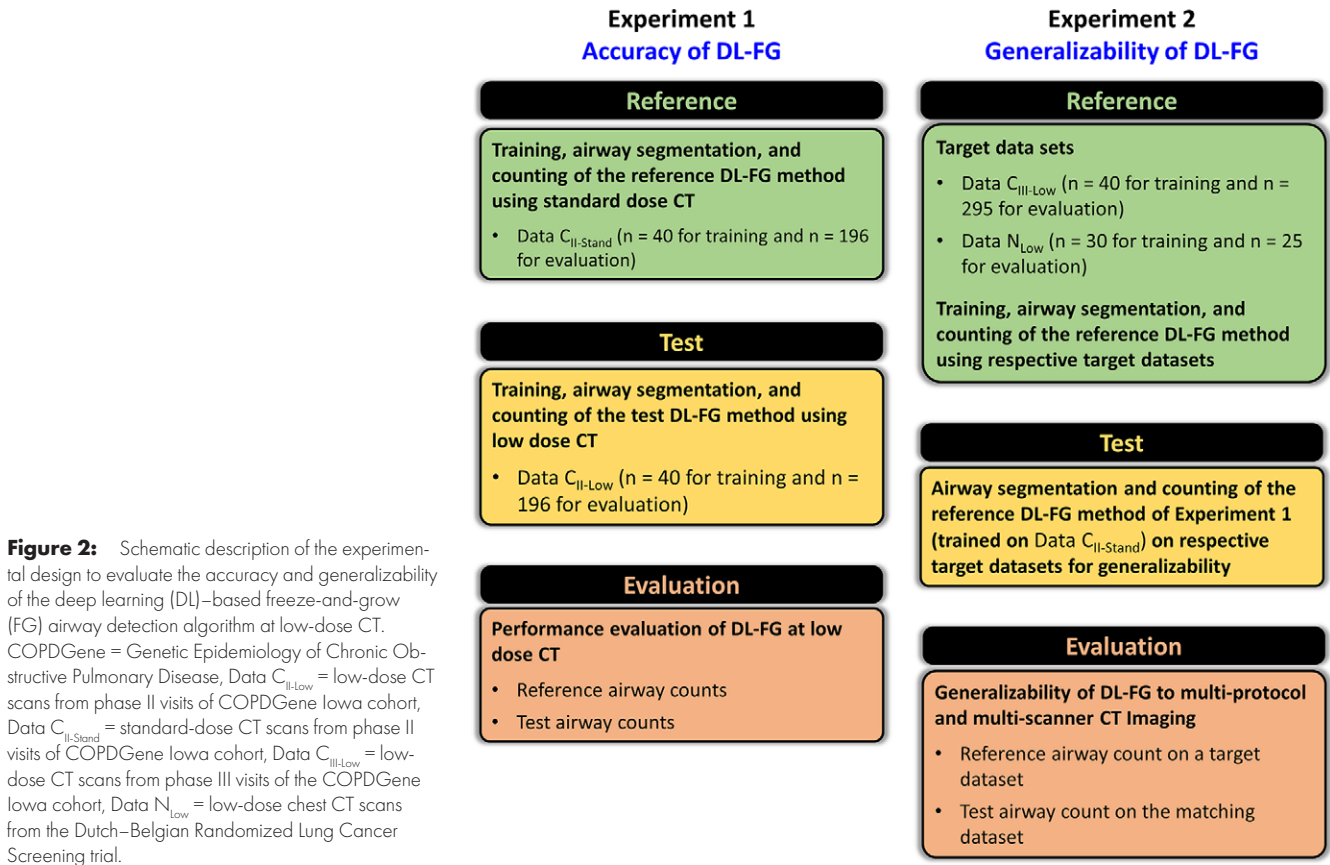
**Experiment 1: Accuracy at low-dose CT.**— A training subset ( $n = 40$ ) of Data  $C_{II-Stand}$  was used, and reference airway segmentations were derived from remaining scans ( $n = 196$ ) in Data  $C_{II-Stand}$ . The motivation was to generate reference segmentations from the best-quality images by using the best-fit network. Test airway segmentations were generated from

matching low-dose scans ( $n = 196$ ) from Data  $C_{II-Low}$  on a DL-FG algorithm trained on remaining scans ( $n = 40$ ) in Data  $C_{II-Low}$ . The goal was to generate low-dose segmentations by using the best-fit network.

**Experiment 2: Multiprotocol and multiscanner generalizability at low-dose CT.**— Two different target data sets, Data  $C_{III-Low}$  and Data  $N_{Low}$ , were used. For a given target data set, a reference DL network was obtained through training on a subset of the target data set and was used to assess the performance of the DL network previously trained on Data  $C_{II-Stand}$  in experiment 1 to examine the generalizability of DL at low-dose CT. The target data set, excluding the training subset, was separately processed through the reference DL network and the previously trained network for evaluation. Forty scans from Data  $C_{III-Low}$  or 30 scans from Data  $N_{Low}$  were used for training, and the remaining 295 or 25 scans, respectively, were used for evaluation.

### Statistical Analysis

The performance metric for DL network optimization was defined as the average contrast between the computed lumen likelihood at a reference lumen voxel and at its nearest voxel in the background. For evaluative experiments, the relative performance metric of a test segmentation was defined as the percentage of branches detected in the matching reference airway tree. The total airway count (TAC) was computed as the total number of detected airways, and the peripheral TAC ( $TAC_p$ ) was computed as the number of airways at and beyond the seventh generation. Airway detectability of a reference set was defined as  $\bar{n}_i / 2^i$ , where  $\bar{n}_i$  is the average number of airways detected at  $i$ th airway tree generation. The false-positive rate was defined as the number of voxels in the test segmentation outside the reference segmentation and was expressed as a percentage of the reference voxel count (28). Descriptive statistics, in-



**Figure 2:** Schematic description of the experimental design to evaluate the accuracy and generalizability of the deep learning (DL)-based freeze-and-grow (FG) airway detection algorithm at low-dose CT. COPDGene = Genetic Epidemiology of Chronic Obstructive Pulmonary Disease, Data  $C_{II-Low}$  = low-dose CT scans from phase II visits of COPDGene Iowa cohort, Data  $C_{II-stand}$  = standard-dose CT scans from phase II visits of COPDGene Iowa cohort, Data  $C_{III-Low}$  = low-dose CT scans from phase III visits of the COPDGene Iowa cohort, Data  $N_{Low}$  = low-dose chest CT scans from the Dutch-Belgian Randomized Lung Cancer Screening trial.

cluding the means and SDs of different performance measures, were computed by using MathWorks MATLAB and Microsoft Excel. Finally, a Wilcoxon signed rank test was conducted to compare the performance of different methods. A  $P$  value less than .05 was considered to indicate statistical significance.

## Results

DL network optimization results are presented in Appendix E1 (supplement). Following the observations of Appendix E1, a network width of 80 kernels and a training data set size of 40 scans were selected for all experiments, except for the multiscanner generalizability experiment involving Data  $N_{Low}$  ( $n = 55$ ), in which 30 scans and 25 scans were used for training and evaluation. The demographic characteristics of participants from different data sets are presented in Table 1.

### Experiment 1: Accuracy at Low-Dose CT

Generational counts and detection rates of reference airways at standard-dose CT are presented in Figure 3A and 3B. Examples of representative reference airway trees for those who have never smoked and participants with different levels of COPD severity grouped by male and female participants are displayed in Figure 3C. Figure 4 presents comparative results of airway segmentation at standard-dose (reference) and low-dose CT. As shown in Figure 4A, at each generation between generation 7 and generation 12, low-dose CT airway counts were lower ( $P = .007$  for generation 7 and  $P < .001$  for generations 8–12) than the reference counts at standard-dose CT.

Low-dose CT scans had a relative performance of 12 349 of 13 482 (91.6%) airways, with an average airway count of 63 at the seventh generation, at which the peak count of reference airways was found. Relative performance was reduced to 404 of 623 (64.8%) airways at the 12th generation. Note that, on an average, only 6.7 airways were detected in reference data at and beyond the 12th generation (Fig 4B). Figure 4D shows CT image slices illustrating airways detected at standard-dose CT but missed at low-dose CT.

Results of airway detection at low-dose CT are summarized in Table 2. All airways were detected at low-dose CT at the segmental level along the five standardized segmental airway paths (RB1 [apical segment of the right upper lobe], RB4 [lateral segment of the right middle lobe], RB10 [posterobasal segment of the right lower lobe], LB1 [apicoposterior segment of the left upper lobe], and LB10 [posterobasal segment of the left lower lobe]); see Figure 5 (29). However, a loss in the number of detected airways at low-dose CT was observed at the subsegmental ( $P = .002$ ) and sub-subsegmental ( $P < .001$ ) levels. Losses ( $P < .001$ ) of 6186 of 62 480 (9.9%) airways in the  $TAC$  and 5755 of 37 684 (15.2%) airways in the  $TAC_p$  were observed at low-dose CT compared with the counts at standard-dose CT. Reference  $TAC$  values of  $347 \pm 67$ ,  $342 \pm 67$ ,  $323 \pm 76$ ,  $266 \pm 56$ ,  $233 \pm 47$ , and  $188 \pm 0$  and  $TAC_p$  values of  $205 \pm 63$ ,  $202 \pm 59$ ,  $189 \pm 69$ ,  $141 \pm 50$ ,  $114 \pm 34$ , and  $69 \pm 0$  were achieved for those who have never smoked and participants at Global Initiative for Chronic Obstructive Lung Disease (GOLD) stages 0–4, respectively.

**Table 1: Characteristics of Participants Included in This Study**

Characteristic	Iowa COPDGene Phase II	Iowa COPDGene Phase III	NELSON Trial*
No. of participants	236	335	55
<b>Demographics</b>			
Age (y)	70 ± 9 (51–86)	73 ± 8 (51–88)	— (50–75)
Men	113 (48)	162 (48)	
Women	123 (52)	173 (52)	
Pack-years <sup>†</sup> smoking	41.1 ± 25.8	41.1 ± 28.0	
Those who have never smoked	59 (25)	50 (15)	
Those who formerly smoked	156 (66)	249 (74)	
Those who smoke	21 (9)	36 (11)	
Body mass index (kg/m <sup>2</sup> )	30.0 ± 6.6	29.1 ± 5.8	
<b>Spirometry (postbronchodilator) results</b>			
FEV <sub>1</sub> /FVC	0.75 ± 0.02	0.74 ± 0.02	
FEV <sub>1</sub> (% predicted)	82.9 ± 24.2	89.9 ± 24.4	
<b>GOLD classification</b>			
-1	49 (21)	57 (17)	
0	51 (22)	112 (33)	
1	23 (10)	40 (12)	
2	38 (16)	57 (17)	
3	15 (6)	15 (4)	
4	1 (0.4)	4 (1)	

Note.—Data are presented as means ± SDs, means ± SDs with ranges in parentheses, or counts with percentages in parentheses. COPDGene = Genetic Epidemiology of Chronic Obstructive Pulmonary Disease, FEV<sub>1</sub> = forced expiratory volume in 1 second, FVC = forced vital capacity, GOLD = Global Initiative for Chronic Obstructive Lung Disease, NELSON = Dutch–Belgian Randomized Lung Cancer Screening.

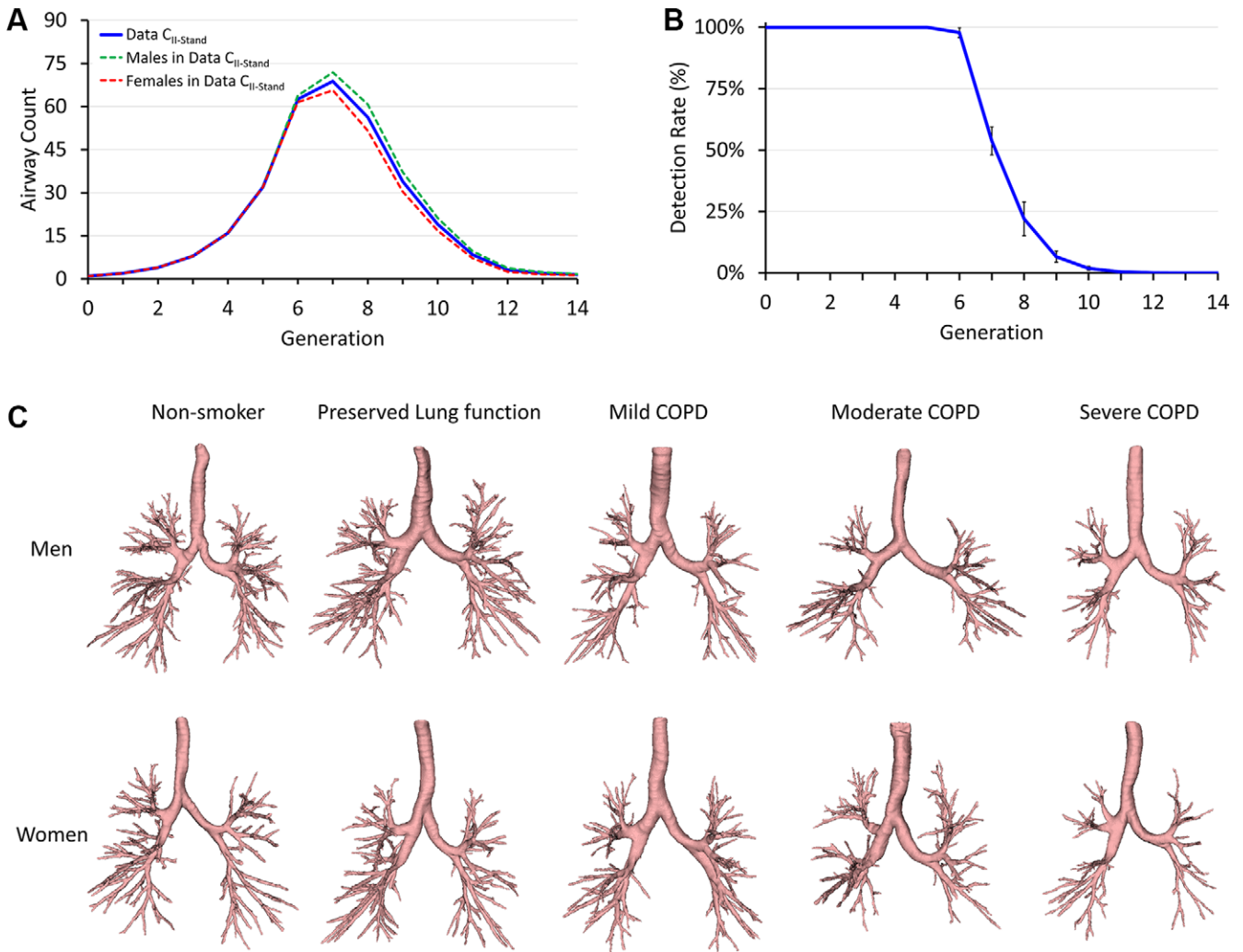
\* Specific demographic information for this data set were not available.

<sup>†</sup> Defined by the American Thoracic Society as the number of packs of cigarettes smoked every day multiplied by the total number of smoking years.

## Experiment 2: Multiprotocol and Multiscanner Generalizability at Low-Dose CT

Generational distributions of reference airway counts for multiprotocol and multiscanner generalizability experiments are presented in Figure 6A; to facilitate comparisons, reference airway counts at standard-dose CT are repeated. Figure 6B presents relative performance for multiprotocol and multiscanner generalizability at low-dose CT. Multiprotocol generalization of DL-FG at low-dose CT caused a loss ( $P = .01$  for generation 6;  $P = .002$  for generation 7; and  $P < .001$  for generations 8–12) in airway counts at each of the sixth to 12th generations. A relative performance of 11 061 of 14 493 (76.3%) was observed at the seventh generation, with a peak generational airway count of 49.1. Relative performance was reduced to 885 of 1861 (47.6%) airways at the 12th generation. Multiscanner generalization at low-dose CT incurred a loss ( $P < .05$ ) of airway counts at each of the seventh to 12th generations. A relative performance of 1326 of 1553 (85.4%) airways was observed at the seventh generation, with a peak generational airway count of 62.1. Relative performance was reduced to 34 of 47 (72.3%) airways at the 12th generation.

The results of multiprotocol and multiscanner generalizability at low-dose CT are presented in Tables 3 and 4. Multiprotocol generalization of DL-FG at low-dose CT caused the loss ( $P = .50$ ) of airways at the segmental level but caused significant losses ( $P < .001$ ) at the subsegmental and sub-subsegmental levels. Multiscanner generalization incurred no significant losses ( $P > .05$ ) of airways at the segmental, subsegmental, or sub-subsegmental levels. Losses ( $P < .001$ ) of 14 526 of 76 453 (19.0%) and 884 of 6908 (12.8%) airways in the TAC and 12 256 of 43 462 (28.2%) and 699 of 3882 (18.0%) airways in the TAC<sub>p</sub> were observed for the multiprotocol and multiscanner generalization, respectively, at low-dose CT. Multiprotocol generalization at low-dose CT achieved TAC values of 221 ± 23, 231 ± 28, 204 ± 30, and 181 ± 26 and TAC<sub>p</sub> values of 93 ± 12, 101 ± 18, 77 ± 15, and 63 ± 14 for those who have never smoked, participants at GOLD stage 0, participants at GOLD stage 1, and participants at GOLD stage 2, respectively. For the multiscanner generalizability experiment, only those who smoke but are without COPD (GOLD stage 0) were studied, and TAC and TAC<sub>p</sub> values of 242 ± 42 and 127 ± 23 were observed for the generalized network at low-dose CT.



**Figure 3:** Reference airway segmentation results on standard-dose total-lung-capacity CT data from phase II visits of the Genetic Epidemiology Chronic Obstructive Pulmonary Disease Iowa cohort (Data  $C_{II-stand}$ ). **(A)** Mean airway counts at different airway tree generations. **(B)** Mean and SD of airway detection rates at different generations. **(C)** Representative reference airways for different groups segmented at standard radiation. The representative participant for each group was selected as the one with the total airway count closest to the respective group mean. COPD = chronic obstructive pulmonary disease.

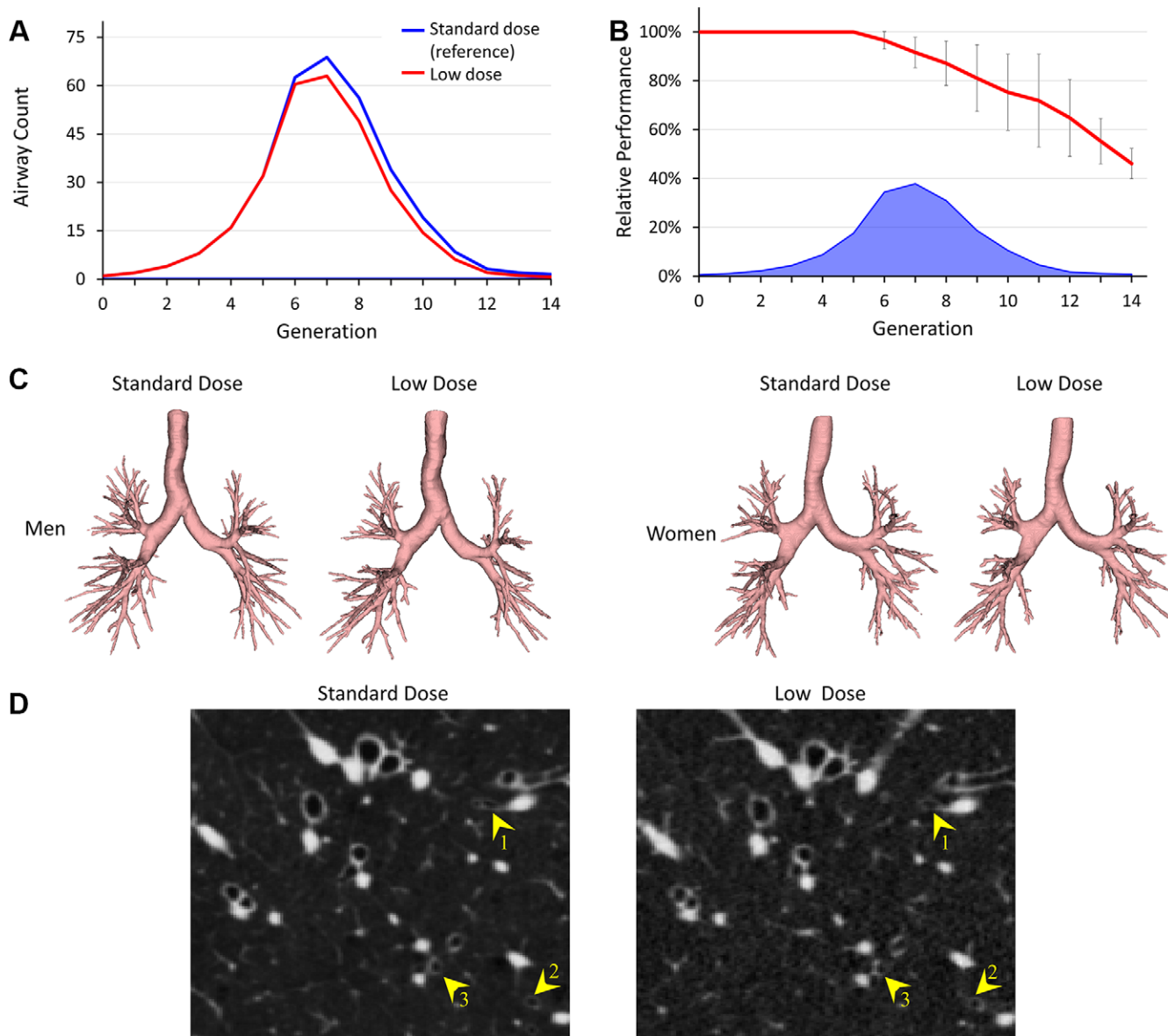
A mean  $\pm$  SD false-positive rate of  $0.3\% \pm 0.2$  with a median of  $0.5\%$  (1972 of 371 825) and range of  $0\%$  (0 of 370 693) to  $1.2\%$  (4871 of 413 312) were observed for multiprotocol generalizability, whereas the false-positive rates for multiscanner generalizability were  $0.3\% \pm 0.3$ ,  $0.3\%$  (1059 of 379 277), and  $0\%$  (0 of 398 174) to  $0.8\%$  (2764 of 346 594), respectively. The means  $\pm$  SDs and 95% CIs of the distances (in voxels) of false-positive voxels and the reference airway segmentation were  $1.5 \pm 0.7$  and (1.4, 1.5), respectively, for multiprotocol generalizability and  $1.3 \pm 0.7$  and (1.0, 1.5), respectively, for multiscanner generalizability. False-positive voxels were primarily located around the reference airway lumen surface, and no false branches were observed in these experiments; see Figure E2 (supplement).

## Discussion

Accurate airway tree segmentation and branch-level tracking at low-dose CT is critical in studying early-stage COPD in multisite longitudinal lung studies (9–12). Results of the accuracy-gauging experiments in this study suggest that at low-dose CT,

compared with standard-dose CT, there are significant losses in detected airways beyond the segmental level, particularly when the same DL-FG method is applied to both standard- and low-dose scans. Losses in detected airways at low-dose CT are consistent with the increase of noise and degradation of image quality and detectable features on low-dose scans (see Fig 4D) (16–18). Although the airway detection losses at low-dose CT were significant, more than 95% of subsegmental and 90% of sub-subsegmental and total reference airways found at standard-dose CT were successfully detected. Approximately 85% of reference peripheral airways were successfully detected at low-dose CT (see Table 2).

Moreover, the results at low-dose CT observed in this study show improvements over prior results at standard-dose CT (10). For example, Kirby et al (30) applied a then state-of-the-art semiautomated method (10) on standard-dose TLC CT scans of those who have never smoked and participants at GOLD stages 0, 1 and 2 from the Canadian Cohort of Obstructive Lung Disease (12). They observed TAC values of 221, 217, 190, and 152



**Figure 4:** Comparative results of airway detection at standard-dose (reference) and low-dose CT imaging. **(A)** Generational distribution of mean airway counts at standard-dose (blue) and low-dose (red) CT. **(B)** Mean and SD of the relative performance at different airway generations at low-dose CT (red). Mean reference airway counts at standard-dose CT are shown at the bottom (blue). **(C)** Representative examples of segmented airways at standard- and low-dose CT for men and women. **(D)** CT image slices show examples of airways (yellow arrowheads) that are detected at standard-dose CT but missed at low-dose CT. CT display settings: level,  $-450$  HU; window,  $1200$  HU.

and  $TAC_p$  values (estimated from figure 2 from Kirby et al [30]) of 80, 76, 57, and 34 for those who have never smoked and participants at GOLD stages 0, 1, and 2, respectively.  $TAC$  and  $TAC_p$  results observed in this study at low-dose CT show considerable improvements over those Kirby et al (30) reported for each of the never-smoking, GOLD 0, GOLD 1, and GOLD 2 populations. These observations suggest that when using the fully automated DL-FG method, low-dose CT imaging provides a viable solution for investigating small peripheral airway phenotypes in early COPD and other lung diseases. It is clinically known that small peripheral airways are the first to be affected in early-onset COPD (6,7), and the detectability of peripheral airways is an important indicator of early-onset COPD in CT-based studies (30,31). Given the increased use of low-dose CT

imaging aimed at reducing cumulative radiation exposure, the observations of this study suggest that DL-FG will add to the sensitivity of early assessment of COPD occurrence and progression and increase the power of studies exploring the changes in airway structure and physiologic function of different lung diseases. Recent guidelines from the U.S. Preventive Services Task Force (32) on CT lung cancer screening (33,34) studies will potentially create a large data set of low-dose chest CT scans, and the current method may provide a suitable tool for quantitative assessment of airway phenotypes.

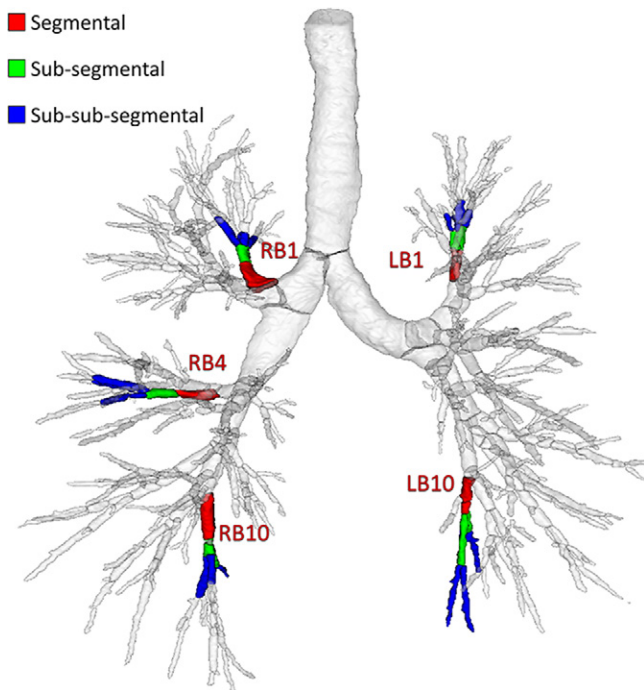
It was observed from the generalizability experiments that the performance of the generalized method was considerably higher in the multiscanner experiment than in the multiprotocol experiment, despite the low-dose and reference standard-dose

**Table 2: Comparative Results of Airway Branch Counts Using DL-FG Airway Segmentation Method at Standard- and Low-Dose CT**

Parameter	Standard	Low	Agreement (%)	Difference (%)	P Value*
Segmental	4.9 ± 0.5	4.9 ± 0.5	100.0	0.0	>.99
Subsegmental	9.8 ± 1.4	9.3 ± 2.5	95.1	4.9	.002
Sub-subsegmental	19.4 ± 3.3	17.4 ± 5.0	90.1	9.9	<.001
TAC	318.8 ± 54.7	287.4 ± 42.2	90.1	9.9	<.001
TAC <sub>p</sub>	193.2 ± 41.2	163.9 ± 30.2	84.8	15.2	<.001

Note.—Data are presented as means ± SDs or percentages. Segmental, subsegmental, and sub-subsegmental airways were counted along the five anatomic paths: RB1 (apical segment of right upper lobe), RB4 (lateral segment of right middle lobe), RB10 (posterobasal segment of right lower lobe), LB1 (apicoposterior segment of left upper lobe), and LB10 (posterobasal segment of left lower lobe). Standard- and low-dose results were obtained by using DL-FG methods trained on standard- and low-dose CT scans, respectively. These results were obtained by using standard- and low-dose CT scans from Genetic Epidemiology of Chronic Obstructive Pulmonary Disease phase II visits at the University of Iowa. DL = deep learning, FG = freeze and grow, TAC = total airway count, TAC<sub>p</sub> = peripheral TAC.

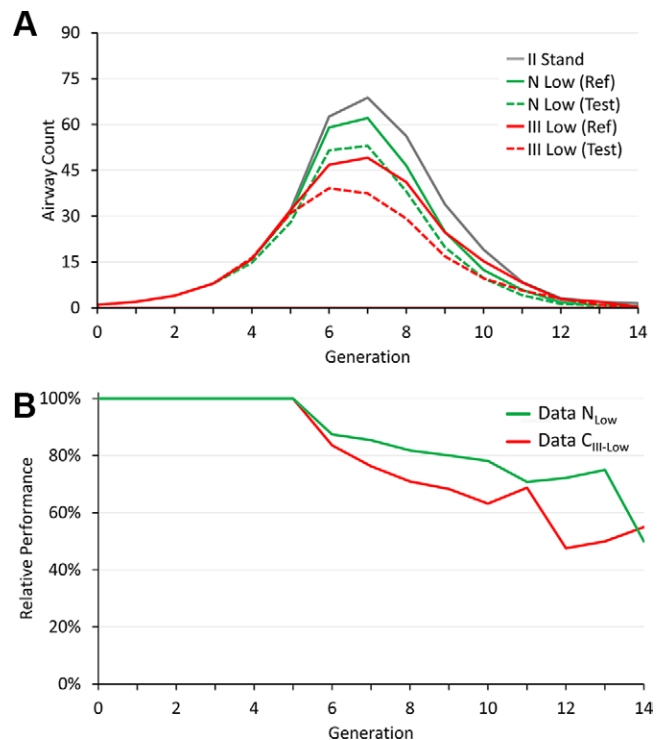
\* Results from Wilcoxon signed rank test.



**Figure 5:** Illustration of segmental, subsegmental, and sub-subsegmental generations along the five standardized bronchopulmonary paths: RB1 (apical segment of right upper lobe), RB4 (lateral segment of right middle lobe), RB10 (posterobasal segment of right lower lobe), LB1 (apicoposterior segment of left upper lobe), and LB10 (posterobasal segment of left lower lobe).

scans being acquired from the same scanner. The primary reason behind this observation may be the population difference in the two experiments, as the participants for the multiscanner experiments were all persons who smoke with preserved lung function. The performance metrics for the two experiments in terms of TAC and TAC<sub>p</sub> were similar, and the gap was considerably reduced when the results of persons who smoke at GOLD stage 0 from the multiprotocol experiment were compared with the observations from the multiscanner experiment.

A common concern with DL-based methods is the cost and time of retraining for a new data set, which can be broken into



**Figure 6:** (A) Generational airway counts for reference (solid) and test (dotted) airway segmentation for multiprotocol (red) and multiscanner (green) generalizability experiments together with reference counts at standard-dose CT (gray line) for the accuracy experiment. (B) Relative performance for the multiprotocol (red) and multiscanner (green) generalizability experiments at low-dose CT. II Stand = the method trained and evaluated on standard-dose CT scans from phase II visits of the Genetic Epidemiology Chronic Obstructive Pulmonary Disease (COPDGene) Iowa cohort (CII-Stand); N Low (Ref) = the method trained and evaluated on low-dose chest CT scans from the Dutch-Belgian Randomized Lung Cancer Screening trial (NLow); N Low (Test) = the method trained on CII-Stand and evaluated on NLow; III Low (Ref) = the method trained and evaluated on low-dose chest CT scans from phase III visits of the COPDGene Iowa cohort (CIII-Low); III Low (Test) = the method trained on CII-Stand and evaluated on CIII-Low; Data N<sub>Low</sub> = the method trained on CII-Stand and evaluated on NLow, Data C<sub>III-Low</sub> = the method trained on CII-Stand and evaluated on CIII-Low.



**Table 3: Results of Multiprotocol Generalizability of DL-FG Airway Segmentation Method at Low-Dose CT**

Parameter	Reference Counts	Generalized Counts	Agreement (%)	Difference (%)	P Value*
Segmental	4.9 ± 0.6	4.9 ± 0.7	99.3	0.7	.50
Subsegmental	9.2 ± 2.7	8.5 ± 3.5	93.4	6.6	<.001
Sub-subsegmental	17.2 ± 7.0	15.2 ± 7.9	90.2	9.8	<.001
TAC	253.7 ± 51.9	205.4 ± 42.0	81.0	19.0	<.001
TAC <sub>p</sub>	143.8 ± 36.3	103.2 ± 23.3	71.8	28.2	<.001

Note.—Data are presented as means ± SDs or percentages. Segmental, subsegmental, and sub-subsegmental airways were counted along the five anatomic paths: RB1 (apical segment of right upper lobe), RB4 (lateral segment of right middle lobe), RB10 (posterobasal segment of right lower lobe), LB1 (apicoposterior segment of left upper lobe), and LB10 (posterobasal segment of left lower lobe). Reference counts were obtained by training a DL-FG method on low-dose scans and then applying the method to low-dose scans, whereas generalized counts were obtained by training a DL-FG method on standard-dose scans and then applying the method to low-dose scans. Standard- and low-dose CT scans used for these results were from Genetic Epidemiology of Chronic Obstructive Pulmonary Disease phase II and III visits at the University of Iowa. DL = deep learning, FG = freeze and grow, TAC = total airway count, TAC<sub>p</sub> = peripheral TAC.

\* Results from Wilcoxon signed rank test.

**Table 4: Results of Multiscanner Generalizability of DL-FG Airway Segmentation Method at Low-Dose CT**

Parameter	Reference Counts	Generalized Counts	Agreement (%)	Difference (%)	P Value*
Segmental	5.0 ± 0.0	5.0 ± 0.0	100.0	0.0	>.99
Subsegmental	10.0 ± 0.0	9.8 ± 1.0	98.0	2.0	.25
Sub-subsegmental	18.6 ± 3.3	18.4 ± 3.7	99.0	1.0	.13
TAC	277.3 ± 42.0	241.9 ± 39.9	87.2	12.8	<.001
TAC <sub>p</sub>	155.3 ± 21.90	127.3 ± 26.3	82.0	18.0	<.001

Note.—Data are presented as means ± SDs or percentages. Segmental, subsegmental, and sub-subsegmental airways were counted along the five anatomic paths: RB1 (apical segment of right upper lobe), RB4 (lateral segment of right middle lobe), RB10 (posterobasal segment of right lower lobe), LB1 (apicoposterior segment of left upper lobe), and LB10 (posterobasal segment of left lower lobe). Reference counts were obtained by training and applying a DL-FG method on low-dose scans from the Dutch–Belgian Randomized Lung Cancer Screening trial, whereas generalized counts were obtained by training a DL-FG method on standard-dose scans from Genetic Epidemiology of Chronic Obstructive Pulmonary Disease phase II visits at the University of Iowa and then applying the method to low-dose Dutch–Belgian Randomized Lung Cancer Screening trial data. DL = deep learning, FG = freeze and grow, TAC = total airway count, TAC<sub>p</sub> = peripheral TAC.

\* Results from Wilcoxon signed rank test.

two components: the preparation time and cost of the training data set and the computational training time. It was observed in this study that approximately 7 days of computation time and 10 hours of expert time were required to prepare a training data set ( $n = 40$ ). Observations from the generalizability experiments suggest there is significant loss in the TAC, TAC<sub>p</sub>, and detected airways beyond the segmental level when a DL-FG algorithm trained on standard-dose scans is applied to multiprotocol and multiscanner CT images. Despite the performance loss that occurs when using a generalized DL-FG algorithm at low-dose CT, it excelled the performance reported by Kirby et al (30), who used the then state-of-the-art semiautomated method at standard-dose CT. Based on the low cost of training and observed performance loss compared with the best achievable results at low-dose CT, we recommend using a DL-FG algorithm after retraining on a target data set.

This study had a few limitations. First, because most participants smoked, the findings are not directly applicable to those who do not smoke. Additionally, the majority of participants

were non-Hispanic White individuals, limiting generalizability of the results to other ethnicities and races. The age distribution of study participants also limits the external validity in younger groups. Comparative results referencing the results of Kirby et al (30) should be interpreted with the caveat that different data sets were used. Finally, this research is at an early stage and requires rigorous investigations evaluating its significance and suitability for adoption in a clinical setting.

In summary, this study validates an automated airway detection algorithm at low-dose CT that uses DL and multiparametric FG methods that is suitable for large CT-based lung studies, which will provide a powerful tool for computing and analyzing airway phenotypes to further the understanding of the mechanisms of disease, severity, and progression.

**Data sharing:** Data generated or analyzed during the study are available from the corresponding author by request.

**Author contributions:** Guarantors of integrity of entire study, E.A.H., P.K.S.; study concepts/study design or data acquisition or data analysis/interpretation, all

authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, all authors; statistical analysis, S.A.N., P.K.S.; and manuscript editing, all authors

**Disclosures of conflicts of interest:** S.A.N. Grants for National Institutes of Health (NIH R01 HL142042, NIH 5U01 HL089897, and NIH S10 OD018526); U.S. patent #10909684, Inventors: P.K.S., E.A.H., S.A.N. Systems and methods for airway tree segmentation. A.P.C. NIH grants (NHLBI, NIEHS, and NCATS); consultant fees from GSK, AstraZeneca, Eli Lilly, RespiaraLabs; nonpaid consultant for VIDA Diagnostics. E.A.H. NIH grants to University of Iowa; royalties or licenses from VIDA Diagnostics paid to University of Iowa; support for attending meetings and/or travel from NIH paid to University of Iowa; founder and shareholder in VIDA Diagnostics, a company commercializing lung image analysis software developed, in part, at the University of Iowa; research-dedicated Siemens Force scanner was purchased through a high-end S10 shared instrumentation grant from the NIH (funding to the University of Iowa); unpaid member of Siemens Photon Counting CT advisory board. P.K.S. Grants for National Institutes of Health (NIH R01 HL142042, NIH 5U01 HL089897, and NIH S10 OD018526); U.S. patent #10909684, Inventors: P.K.S., E.A.H., S.A.N. Systems and methods for airway tree segmentation, year 2021.

## References

- Ahmad FB, Anderson RN. The leading causes of death in the US for 2020. *JAMA* 2021;325(18):1829–1830.
- Kiley JP, Gibbons GH. COPD National Action Plan: addressing a public health need together. *Chest* 2017;152(4):698–699.
- Han MK, Martinez CH, Au DH, et al. Meeting the challenge of COPD care delivery in the USA: a multiprovider perspective. *Lancet Respir Med* 2016;4(6):473–526.
- Jeffery PK. Structural and inflammatory changes in COPD: a comparison with asthma. *Thorax* 1998;53(2):129–136.
- Jeffery PK. Comparison of the structural and inflammatory features of COPD and asthma. Giles F. Filley lecture. *Chest* 2000;117(5 Suppl 1):251S–260S.
- Bignon J, Khoury F, Even P, Andre J, Brouet G. Morphometric study in chronic obstructive bronchopulmonary disease. Pathologic, clinical, and physiologic correlations. *Am Rev Respir Dis* 1969;99(5):669–695.
- Hogg JC, Macklem PT, Thurlbeck WM. Site and nature of airway obstruction in chronic obstructive lung disease. *N Engl J Med* 1968;278(25):1355–1360.
- Smith BM, Traboulsi H, Austin JHM, et al; MESA Lung and SPIROMICS Investigators. Human airway branch variation and chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A* 2018;115(5):E974–E981.
- Regan EA, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) study design. *COPD* 2010;7(1):32–43.
- Sieren JP, Newell JD Jr, Barr RG, et al; SPIROMICS Research Group. SPIROMICS protocol for multicenter quantitative computed tomography to phenotype the lungs. *Am J Respir Crit Care Med* 2016;194(7):794–806.
- Couper D, LaVange LM, Han M, et al; SPIROMICS Research Group. Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* 2014;69(5):491–494.
- Bourbeau J, Tan WC, Benedetti A, et al; CanCOLD Study Group. Canadian Cohort Obstructive Lung Disease (CanCOLD): fulfilling the need for longitudinal observational studies in COPD. *COPD* 2014;11(2):125–132.
- Donohue KM, Hoffman EA, Baumhauer H, et al. Cigarette smoking and airway wall thickness on CT scan in a multi-ethnic cohort: the MESA Lung Study. *Respir Med* 2012;106(12):1655–1664.
- van Ginneken B, Baggerman W, van Rikxoort EM. Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Metaxas D, Axel L, Fichtinger G, Székely G, eds. *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. Berlin, Germany: Springer, 2008; 219–226.
- Aberle DR, Adams AM, et al; National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365(5):395–409.
- Goldman LW. Principles of CT: radiation dose and image quality. *J Nucl Med Technol* 2007;35(4):213–225; quiz 226–228.
- Hammond E, Sloan C, Newell JD Jr, et al. Comparison of low- and ultralow-dose computed tomography protocols for quantitative lung and airway assessment. *Med Phys* 2017;44(9):4747–4757.
- Tschirren J, Hoffman EA, McLennan G, Sonka M. Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. *IEEE Trans Med Imaging* 2005;24(12):1529–1539.
- Nadeem SA, Hoffman EA, Sieren JC, et al. A CT-based automated algorithm for airway segmentation using freeze-and-grow propagation and deep learning. *IEEE Trans Med Imaging* 2021;40(1):405–418.
- Ru Zhao Y, Xie X, de Koning HJ, Mali WP, Vliegenthart R, Oudkerk M. NELSON lung cancer screening study. *Cancer Imaging* 2011;11 Spec No A(1A):S79–S84.
- van Ginneken B, Armato SG 3rd, de Hoop B, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Med Image Anal* 2010;14(6):707–722.
- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W, eds. *Proceedings of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Vol 9901, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2016; 424–432.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Proceedings Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Vol 9351, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2015; 234–241.
- Jin D, Iyer KS, Chen C, Hoffman EA, Saha PK. A robust and efficient curve skeletonization algorithm for tree-like objects using minimum cost paths. *Pattern Recognit Lett* 2016;76:32–40.
- Saha PK, Borgefors G, Sanniti di Baja G. A survey on skeletonization algorithms and their applications. *Pattern Recognit Lett* 2016;76:3–12.
- Nadeem SA, Hoffman EA, Sieren JP, Saha PK. Topological leakage detection and freeze-and-grow propagation for improved CT-based airway segmentation. In: Angelini ED, Landman BA, eds. *Proceedings of SPIE: medical imaging 2018—image processing*. Vol 10574. Bellingham, Wash: International Society for Optics and Photonics, 2018; 105741A.
- Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31(3):1116–1128.
- Garcia-Uceda A, Selvan R, Saghir Z, Tiddens HAWM, de Bruijne M. Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks. *Sci Rep* 2021;11(1):16001.
- Boyden EA. Segmental anatomy of the lungs: a study of the patterns of the segmental bronchi and related pulmonary vessels. New York, NY: Blakiston Division, McGraw-Hill, 1955; 185–200.
- Kirby M, Tanabe N, Tan WC, et al; CanCOLD Collaborative Research Group, the Canadian Respiratory Research Network. 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* 2018;197(1):56–65.
- Diaz AA, Valim C, Yamashiro T, et al. Airway count and emphysema assessed by chest CT imaging predicts clinical outcome in smokers. *Chest* 2010;138(4):880–887.
- Potter AL, Bajaj SS, Yang CJ. The 2021 USPSTF lung cancer screening guidelines: a new frontier. *Lancet Respir Med* 2021;9(7):689–691.
- Hoffman RM, Sanchez R. Lung cancer screening. *Med Clin North Am* 2017;101(4):769–785.
- Houston T. Screening for lung cancer. *Med Clin North Am* 2020;104(6):1037–1050.