

Geodesic atlas-based labeling of anatomical trees: Application and evaluation on airways extracted from CT

Aasa Feragen[†], Jens Petersen[†], Megan Owen, Pechin Lo, Laura Hohwü Thomsen, Mathilde Marie Winkler Wille, Asger Dirksen, Marleen de Bruijne

Abstract—We present a fast and robust atlas-based algorithm for labeling airway trees, using geodesic distances in a geometric tree-space. Possible branch label configurations for an unlabeled airway tree are evaluated using distances to a training set of labeled airway trees. In tree-space, airway tree topology and geometry change continuously, giving a natural automatic handling of anatomical differences and noise. A hierarchical approach makes the algorithm efficient, assigning labels from the trachea and downwards. Only the airway centerline tree is used, which is relatively unaffected by pathology. The algorithm is evaluated on 80 segmented airway trees from 40 subjects at two time points, labeled by 3 medical experts each, testing accuracy, reproducibility and robustness in patients with Chronic Obstructive Pulmonary Disease (COPD). The accuracy of the algorithm is statistically similar to that of the experts and not significantly correlated with COPD severity. The reproducibility of the algorithm is significantly better than that of the experts, and negatively correlated with COPD severity. Evaluation of the algorithm on a longitudinal set of 8724 trees from a lung cancer screening trial shows that the algorithm can be used in large scale studies with high reproducibility, and that the negative correlation of reproducibility with COPD severity can be explained by missing branches, for instance due to segmentation problems in COPD patients. We conclude that the algorithm is robust to COPD severity given equally complete airway trees, and comparable in performance to that of experts in pulmonary medicine, emphasizing the suitability of the labeling algorithm for clinical use.



1 INTRODUCTION

Computed Tomography (CT) is an important tool in the analysis of diseases affecting pulmonary airways. Using image segmentation methods, three-dimensional models of the airway surfaces can be constructed, and their dimensions measured. Measurements such as lumen diameter, airway wall thickness, and bifurcation angle are, however, dependent on the location in which they are made; e.g., it follows from [1] that the classification

boundary and accuracy of using airway wall thickness for COPD prediction is different in different locations in the lung. As a consequence, it is crucial to determine anatomically corresponding positions in different airway trees in order to robustly compare measurements across patients [2]. One way to solve this problem is to identify the airway tree branches by their anatomical names, and compare measurements in identically named airway branches [1], [3], [4]. Identifying the anatomical names is nontrivial, since the topology of the airway tree changes from person to person, and the segmented trees have additional differences introduced by noise, including missing and spurious branches.

Several airway branch labeling algorithms have appeared in the literature. Mori et al [5], [6] initially use rule-based assignment of a pre-determined labeled tree topology, later with multiple pre-determined topology options [7]. Gu et al [8] label the lobe branches by assuming a fixed tree topology and assigning labels based on the (x, y, z) coordinates of bifurcations. These methods are sensitive to topological variation in populations of airway trees. Van Ginneken et al. [9], Mori et al [10] and Lo et al. [11] assign branch labels using supervised learning on branch features, in some cases [9], [11] constrained by assumptions on airway tree topology. Among the features used are branch length, radius, orientation, cross-sectional shape and bifurcation angle. Branch radius is sensitive to diseases like asthma, cystic fibrosis, tuberculosis and Chronic Obstructive Pulmonary Disease (COPD) [3], [12]–[14]. Moreover, airway branch length, shape and bifurcation angle are sensitive

- [†] A. Feragen and J. Petersen contributed equally.
- A. Feragen and J. Petersen are with the Department of Computer Science, University of Copenhagen, Denmark.
- M. Owen is with the Department of Mathematics and Computer Science, Lehman College, City University of New York, USA
- P. Lo is with the Center for Computer Vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine, University of California, Los Angeles, USA.
- L.H. Thomsen, M.M.W. Wille and A. Dirksen are with Lungemedicinsk Afdeling, Gentofte Hospital, Denmark.
- M. de Bruijne is with the Department of Computer Science, University of Copenhagen, Denmark and with the Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC - University Medical Center Rotterdam, The Netherlands.
- This paper has a supplementary graphical abstract available at <http://ieeexplore.ieee.org>, provided by the authors. This includes an MP4 format movie clip illustrating the main steps of the labeling algorithm and the main results of the paper. This material is 16.8 MB in size. The supplementary material can also be downloaded from <http://image.diku.dk/aasa/GraphicalAbstract.tar.gz>.
- Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.
E-mail: {aasa,phup,pechin,marleen}@diku.dk,
megan.owen@lehman.cuny.edu, laura.hohwue.thomsen@regionh.dk,
MWIN0030@geh.regionh.dk, adi@dadlnet.dk

to anatomical differences in topology and, in particular, missing branches in the airway segmentation. For instance, if only one branch in a bifurcation is detected, the result will be a longer branch with different shape and different endpoint bifurcation angles. It is known that fewer branches are typically detected in airway trees from subjects with COPD [4], [12], leading to topological irregularities. These segmentation problems will affect any method which enforces constraints on airway tree topology.

Anatomical tree labeling is closely related with anatomical tree matching, or the problem of matching the branches or bifurcations of one tree to those of another, in the sense that matching an unlabeled tree to a labeled one will generate a label transfer to the unlabeled tree. Pisupati et al [15] use tree matching in airway trees of dogs. Graham and Higgins [16] use a dynamical programming approach to graph matching for matching pairs of airway trees. Tschirren et al [17] and Kitaoka et al [18] label airway trees using association graphs for pairs of trees, which incorporate information from both trees, such that maximal cliques in the association graph induce branch matchings between the original graphs. A similar approach is used by Metzen et al [19] for matching both airway trees and vessel trees in the liver, as well as by Bogunovic et al [20] for labeling the Circle of Willis. While branch shape features go into the construction of the association graph, the possible branch matches are subject to strict constraints as the matching is equivalent to identifying maximal isomorphic subtrees. In particular, the association graph model is not able to take into account the way that arbitrarily small changes in geometric branch features (e.g. branch length) can lead to new topologies, as in Fig. 1.

Other, more geometric approaches, also appear: Smeets et al. [21] match branches from lung vessel trees using pairwise distances between bifurcations both in 3D Euclidean space and along the tree to generate distance matrix "fingerprints", which are matched. Bülow et al. [22] match airway tree branches without connectivity information, using only branch shape context. Kaftan et al. [23], match tree *paths* rather than branches, thus avoiding the difficulty with different tree-topological structures, but also losing all information stored in the topological structure. In particular, this model does not generate branch labels, as the branch division is lost.

Ross et al. [24] employ a probabilistic approach, using a hidden Markov tree model to infer labels of discretely sampled points in the airway tree. Points within individual branches are labeled down to the lobar level after which generations are assigned. The approach is interesting in that it is able to use information on the airway tree structure without being limited by a certain assumed topology.

Feragen et al [25] label airways based on geodesics, or shortest paths, in a space of trees. Their tree-space has no known efficient algorithm for computation of geodesics, making their method too computationally expensive to

label the whole airway tree. In this work we use a more restrictive space of leaf-labeled trees [26], where geodesics can be computed in polynomial time [27].

We present a novel atlas-based algorithm for automatic airway branch labeling, based on geodesic distances in a space of leaf-labeled trees. The labeling algorithm works by suggesting a set of branch label configurations, each forming a suggested labeled airway tree. These suggestions are compared using geodesic distances from each suggested labeled tree to airway trees within a training set labeled by clinical experts, and the optimal suggested labeling is returned. Labels are thus assigned automatically from the trachea and downwards in a hierarchical fashion.

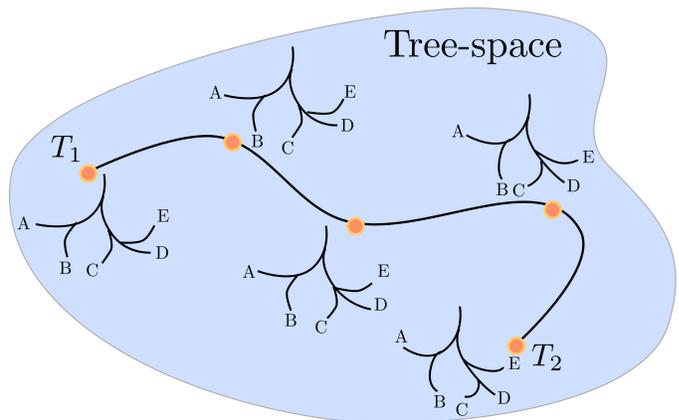


Fig. 1. Since tree-space is a path connected space, any two trees are joined by a path in tree-space, corresponding to a tree deformation along which tree topology and branch geometry changes. The geodesic distance between two trees is the length of the shortest path connecting them. Note the topological transition in the subtree spanned by the leaves C - E .

The tree-space framework is able to compute distances between trees with different topologies, allowing for flexibility in tree topology. This is possible because in tree-space, tree topology and branch geometry are allowed to change continuously. See Fig. 1 for an illustration. From a practical point of view, this allows us to take advantage of the whole training set without being restricted by topological airway tree differences.

The only feature used by the labeling algorithm is the airway centerline tree, divided into branches. The algorithm does not depend directly on the division of the segmented airway tree into branches, but rather on a subtree spanned by the labeled branches, as explained in Section 2.3. This enables the algorithm to tackle structural noise such as false or missing branches, as opposed to methods that work only with the originally segmented branches. The hierarchical implementation makes the algorithm sufficiently fast to be of practical use.

A thorough evaluation of the labeling algorithm is made on a set of 80 segmented airway trees from 40 subjects scanned at two different time-points. Each air-

way tree was labeled by 3 clinical experts. There are subjects without COPD and subjects with different stages of COPD, ranging from mild to severe. The evaluation includes a comparison of accuracy and reproducibility of the algorithm to that of the clinical experts, as well as robustness of accuracy and reproducibility to disease severity. Moreover, reproducibility and robustness of reproducibility to COPD severity is also tested on a large data set from a longitudinal lung cancer screening trial using 8724 CT images from 1900 individuals.

A preliminary version of the work presented here appeared in [28]. In comparison with the earlier paper, changes have been made to the algorithm, making the hierarchy less sensitive to missing RUL branches as detailed in Section 2.5. The exposition has been extended to give a far more comprehensive explanation of the tree-space methodology used. Finally, the experimental validation has been significantly extended. First, our labeled data set has been doubled in size and is now manually annotated by three clinical experts. Second, an evaluation of reproducibility on a large longitudinal study has been conducted, as well as a statistical analysis showing that correlation between increased COPD severity and decreased labeling reproducibility is due to segmentation problems in patients with COPD.

2 METHODS: AIRWAY BRANCH LABELING

The airway branch labels illustrated in Fig. 2 correspond to the division of the lung into compartments. The *Trachea* is the root branch that feeds the lungs. From the trachea, the *left and right main bronchi* (LMB and RMB) lead to the left and right lungs. Beneath the LMB and RMB branches, the *left and right upper lobe* (LUL, RUL), *right middle lobe* (R4+5), and the *left and right lower lobe* (LLB, RLL) bronchi lead to the different lobes. The *Bronchus Intermedius* (BronchInt) feeds the middle and lower lobes in the right lung. Feeding the subdividing segments within the lobes, the *right- and left segment branches numbered 1-10* (R1-R10, L1-L10) lead to the segments, with up to 10 segments in each lung. In the left lung, the branch L4+5 is a counterpart of the right middle lobe branch R4+5. The L4+5 does not lead to a physically separate lobe, but it nevertheless feeds the segments L4-L5. The hierarchical subdivision into two lungs, lobes within the lungs, and segments within the lobes, introduces a natural hierarchy in the labeled branches as well. Due to topological differences and additional/missing branches in different subjects, a number of intermediate branch names appear in the literature, whose presence in a particular anatomical airway tree depends on its topology. If the locations of all segment branch labels are known, along with the airway tree structure, then it is straight-forward to reconstruct the remaining branch labels higher in the hierarchy. In this sense, a *leaf-labeled airway tree*, where the leaf labels are segment labels, is equivalent to a *labeled airway tree*.

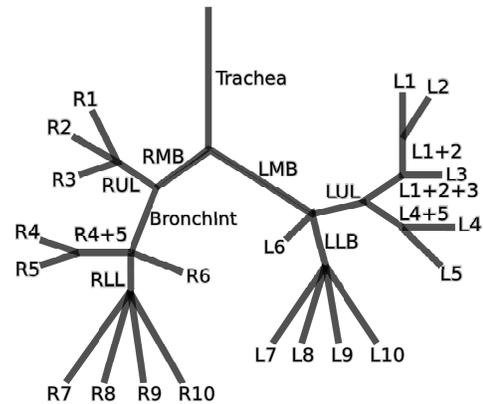


Fig. 2. Airway branch labels used by the labeling algorithm.

2.1 The labeling algorithm: An overview

The input to the labeling algorithm is a connected centerline tree extracted from an airway tree segmentation, divided into branches by bifurcation points. The airway segmentation, centerline extraction and bifurcation detection algorithms used in our experiments are detailed in Section 3.1.

Based on the extracted airway centerlines, each branch is represented by 6 landmark points sampled equidistantly along the branch centerline, translated so that the first landmark point is placed at the origin. Thus, ignoring the first origin landmark point, each branch e is represented by a vector $x_e \in (\mathbb{R}^3)^5 = \mathbb{R}^{15}$. Each airway tree is normalized by the person's height as an isotropic scaling parameter. The person's height was chosen over alternative normalization parameters such as lung volume because height is unaffected by disease.

The general goal of the labeling algorithm is, for an arbitrary unlabeled airway tree T , to optimally assign the set of segment labels $\{L1, \dots, L10, R1, \dots, R10\}$, corresponding to the 20 segment bronchi, to branches in the centerline tree. As outlined in Algorithm 1, the basic labeling algorithm contains a labeling suggestion step and a labeling evaluation step, after which an optimal labeling is selected. In practice, for the sake of computational efficiency, this algorithm is repeated in a hierarchical fashion, assigning labels from the top and downwards. This is detailed in Sec. 2.5.

Algorithm 1 Overview of the basic labeling algorithm

- 1: **Input:** Unlabeled tree T
 - 2: **Input:** Training set of labeled trees $\{T_i | i \in I\}$
 - 3: Generate a set \mathcal{L} of suggested labelings $L \in \mathcal{L}$
 - 4: **for** suggested labelings $L \in \mathcal{L}$ **do**
 - 5: $T_L \leftarrow T$ with suggested labeling L
 - 6: Compute distances $\{d(T_L, T_i) | i \in I\}$
 - 7: Compute $f_L = \sum_{i \in I} d(T_L, T_i)$
 - 8: **end for**
 - 9: **Output:** Labeled tree $T_L = \operatorname{argmin}_{L \in \mathcal{L}} f_L$
-



Fig. 3. From a configuration of leaf labels we extract the subtree spanned by the labels and prune off the rest, obtaining the subtree spanned by the labels, a leaf-labeled tree which can be compared to the training trees.

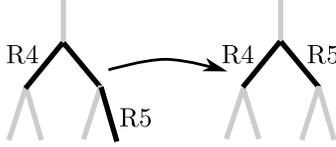


Fig. 4. Each assigned label is backtraced to the branch closest to the root that is not already part of the subtree spanned by the other labels.

Algorithm 1 contains a label suggestion step (line 3) and a label selection step (line 9). In the label suggestion step, a series of potential label configurations are suggested. This is explained in detail in Section 2.5. In the labeling step, the optimal branch label assignment is selected as the configuration that minimizes the sum of distances to manually labeled trees from a training set. The distance used is the geodesic distance in the space of leaf-labeled trees, as detailed in Section 2.4 below.

2.2 Trees

By *tree* we mean a rooted tree, defined as a triple $T = (V, E, r)$ where V is a finite set of vertices, $E \subset V \times V$ is a set of edges so that the corresponding graph is connected and has no cycles, and r is a designated root vertex. Anatomical or biological transportation systems often have a natural source node which can be used as a root. For airway trees, the trachea provides a natural and easy-to-identify root branch [29]. Given any edge $e \in E$, any other edge $\tilde{e} \in E$ which sits on the path through the tree from e to the root is said to be *above* e . If \tilde{e} is above e , then we say that e is *below* \tilde{e} .

A *labeling* of T is a map $L: X \rightarrow E$, which assigns unique labels from a label set X to some but not necessarily all edges. In this paper, we are particularly interested in *leaf-labeled trees*. A *leaf* in T is an edge which does not have any other edges below it. A leaf-labeled tree on the leaf label set X is a tree endowed with a bijective labeling $L: X \rightarrow E_l$, where $E_l \subset E$ are the leaf edges in T . In particular, $|X|$ must equal the number of leaves in T .

2.3 From labeled airway trees to leaf-labeled trees

Given segmented airway trees, we wish to extract leaf-labeled subtrees in such a way that particular sets of branches play the roles of leaves; for instance, the fixed

set of leaf labels $\{L1, \dots, L10, R1, \dots, R10\}$. However, segmented airway trees have variable size and usually, many branches are detected below the segment level. In order to study airway trees using a framework for leaf-labeled trees we define, given any labeling $L: X \rightarrow E$, the *subtree spanned by the labels* as the tree obtained by removing all edges in the tree which are not found on the path from the root to an edge labeled by L , as in Fig. 3. Consecutive edges which are joined by a vertex of order 2 will be concatenated, as is the case with parent branch of $R9$ and $R10$ in Fig. 3. When two edges e_1 and e_2 are concatenated into an edge e , the shape vector x_e will be recomputed from the concatenation of the branch centerlines corresponding to e_1 and e_2 . After labeling, each label is backtraced through the path to the root, as in Fig. 4. We only consider *admissible* label configurations, defined as labelings where the leaf labels will all be attached to *leaves* in the subtree spanned by the labels. This is equivalent to excluding labelings where two leaf labels are assigned to two branches where one branch is above the other.

For a tree T with a labeling $L: X \rightarrow E$, we denote by T_L the subtree spanned by labels assigned by L .

2.4 Tree-space and tree-space distances

The tree-to-tree distances used in this paper are measured in a *tree-space*. This tree-space is a straight-forward generalization of the phylogenetic tree-space defined and studied in [26], where single-dimensional shape vectors on the branches have been generalized to multi-dimensional ones. Below, we give a brief description of tree-space and its properties in order to give the reader intuition for how tree-space works as a model for anatomical trees. For the details of the mathematical framework and the algorithm for computing geodesics, we refer the reader to the original papers [26], [27].

2.4.1 Tree-space

Each point in tree-space is a leaf-labeled tree, with leaves labeled by some fixed set X , for instance $X = \{L1, \dots, L10, R1, \dots, R10\}$. Tree-space is a *path connected space*, which means that any two trees can be joined by a path in tree-space. Moving along such a path corresponds to deforming the trees, as in Fig. 1. Moreover, in tree-space there will always be a unique *shortest* path, called a *geodesic*, joining any given two trees. The length of the geodesic defines a metric distance measure between the two trees [26], called the *geodesic distance*, which will be used by our algorithm.

Each set of trees having a given topology forms a tree-space *orthant*, as illustrated in Figs. 5 and 6. An orthant is a lower-dimensional Euclidean space (or, in the case where edges are described by edge length, a positive

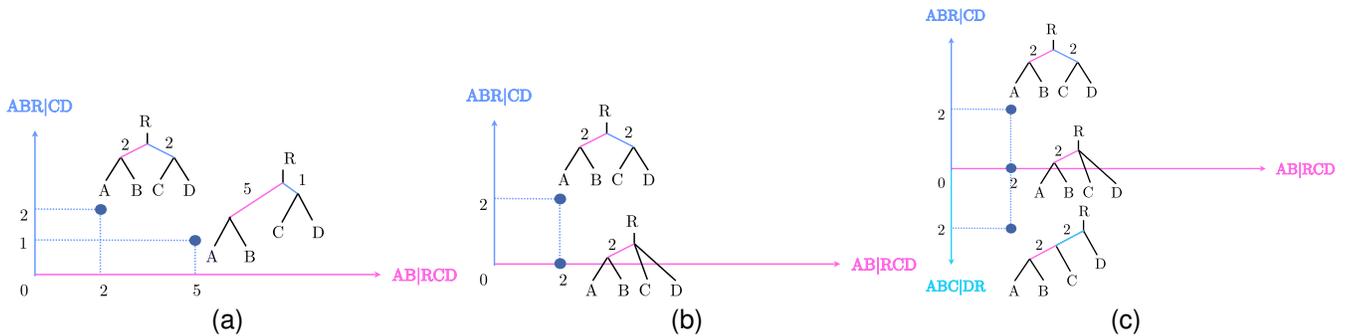


Fig. 5. Tree-space is a union of *orthants*, each corresponding to a specific leaf-labeled tree topology. (a) Different points in the orthant are trees with identical topology but different shapes. (b) Points at the boundary of an orthant are points where one edge is described by a zero vector. Geometrically, that edge has been contracted. (c) Orthants with different tree topologies meet at the boundaries where the contracted edges give rise to new, identical tree topologies.

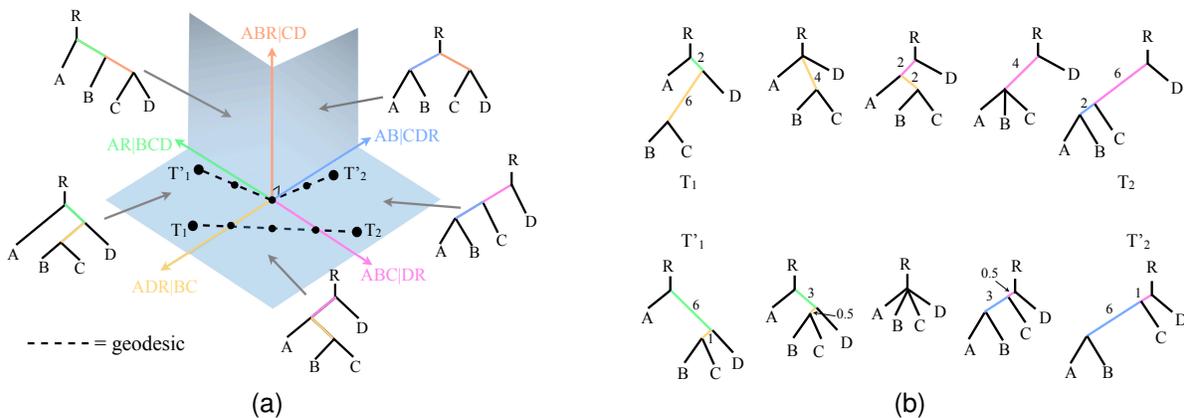


Fig. 6. (a) Tree-space is a proper subset of the Euclidean space $(\mathbb{R}^{15})^S$, and the distance between two trees T_1 and T_2 is the length (measured in the Euclidean space) of the geodesic, or shortest path, in tree-space from T_1 to T_2 . Note that the geodesic from T_1' to T_2' is *not* a straight line, giving different topological transitions throughout the two paths connecting T_1 to T_2 , and T_1' to T_2' . This is illustrated in (b), where trees are sampled along the two geodesic paths. *We illustrate the tree-space using edge length for edges rather than their 3D shape; this is done for illustrative purposes only. The same behavior carries over to edges with shape-vector attributes. Furthermore, the 5 axes depicted above each correspond to their own dimension, and have only been embedded into \mathbb{R}^3 for illustrative purposes.*

orthant of a Euclidean space)¹. Tree-space consists of a set of orthants, one for each possible tree topology, glued together to form a connected space, as in Figs. 5 and 6. Tree-space geodesics are finite concatenations of straight-line segments within single orthants, which are joined at the orthant boundaries. The geodesic path formed by joining the segments can either result in a straight path across the orthant boundary (as for the geodesic from T_1 to T_2 in Fig. 6(a)), or can bend around a tree-space corner (as for the path from T_1' to T_2' in Fig. 6(a)). In general,

1. Formally, an orthant is the part of Euclidean space where all coordinates are nonnegative. When edges are described by their length, as with phylogenetic trees, tree-space orthants are precisely such Euclidean orthants; we call them *positive orthants*. In our paper, branches can have negative 3D coordinates, making tree-space orthants larger than Euclidean orthants. To keep the terminology consistent with phylogenetic tree-space papers, we use the word "orthant" for these larger sub-spaces of tree-space.

geodesics are solutions to the problem

$$\operatorname{argmin} \left\{ \sum_{n=1}^{N-1} \|t_{n+1} - t_n\| \right\} \quad (1)$$

where $T_1 = t_1, T_2 = t_N, t_n$ and t_{n+1} sit on the boundaries of the same orthant, $\|\cdot\|$ is the Euclidean norm within the orthant, and the minimization is over all possible such sets $T_1 = t_1, t_2, \dots, T_N = T_2$.

The algorithm [27] for computing the geodesic distance between two trees works by recursively determining the sequence of orthants containing the geodesic as follows. If the trees are in the same orthant (i.e. if they have the same topology), the algorithm terminates and returns the Euclidean distance between the two trees in the orthant. Otherwise, if the trees are in different orthants, the algorithm starts by computing an initial path connecting the two trees, which goes straight from the first tree, to the origin, and back to the second

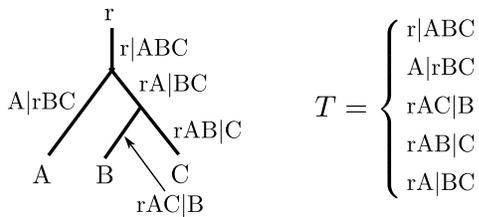


Fig. 7. Tree edges are topologically identified with partitions of the leaf label set X , and a tree topology is characterized uniquely by the partitions that define its edges.

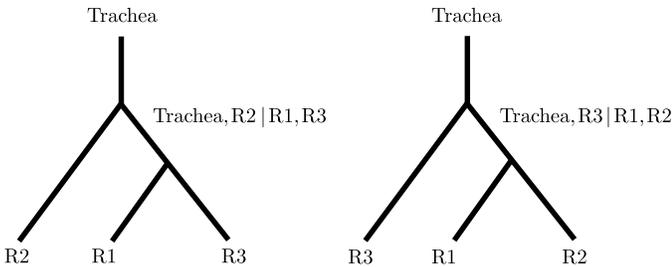


Fig. 8. Certain pairs of label set partitions represent edges that cannot exist in a tree simultaneously. An example for the leaf label set $X = \{R1, R2, R3\}$ is an edge that splits $\{R1, R2\}$ off from the rest of the tree and an edge that splits $\{R1, R3\}$ off.

tree. The algorithm checks if this is the geodesic by looking for an orthant such that modifying the current path to go through this orthant instead of the origin gives a shorter path, i.e., whether this orthant provides a “shortcut” which avoids the origin. If such an orthant exists, we add it to our sequence and calculate the geodesic through the three orthants. Again, we check if this is the overall geodesic by checking each point where the path changes orthants for a new orthant to add into the orthant sequence, such that going through this new orthant will give a shorter path. The new geodesic through the expanded orthant sequence is computed, and this process is repeated until no more orthants can be added. The length of the geodesic through the resulting orthant sequence is the geodesic through tree-space. This algorithm is developed in [27] where details and code can be found.²

2.4.2 Tree-space as a subset of Euclidean space

An alternative way to understand tree-space and its geodesic distance between trees comes from the fact that tree-space is naturally embedded as a subset of a high-dimensional Euclidean space, which can be seen as follows:

Each edge in the leaf-labeled tree can be identified topologically by a partition of X into the leaves below the edge, and the remaining leaves (including the root),

as in Fig. 7. Let S denote the set of all possible partitions of X . Any leaf-labeled tree topology with leaf label set X corresponds to a binary vector $\{0, 1\}^S$, where each coordinate $s \in S$ that is set to 1 indicates that the particular label set bipartition s identifies an edge in the tree topology. Adding shape to the picture, a tree will uniquely correspond to a vector in $(\mathbb{R}^{15})^S$, where each consecutive set of 15 coordinates corresponds to a possible partition s of X . If the edge associated with that partition appears in the tree, then those 15 coordinates will be its branch vector, and otherwise they are all 0.

Certain edges can never appear in a tree together. An example is shown in Fig. 8, where an edge that splits $\{R1, R2\}$ off from the rest of the tree and an edge that splits $\{R1, R3\}$ off from the rest of the tree cannot possibly appear in the same tree. This means that tree-space is not all of $(\mathbb{R}^{15})^S$, but consists precisely of those vectors in $(\mathbb{R}^{15})^S$ that correspond to trees. Thus, tree-space is a proper subset of Euclidean space. The geodesic distance $d(T, T')$ between two trees T and T' defined in [26] coincides with the length of the shortest path between T and T' that remains fully within this restricted subspace, length being measured in the ambient Euclidean space using the Euclidean metric. An analytic formula for this distance d does not exist, but as described above, a recursive algorithm exists and can be computed recursively in polynomial time [27].

2.4.3 Example of tree-space geodesic

To return to a concrete example, consider Fig. 6, where a concrete example is given of two geodesics for which the endpoint trees (T_1, T_2) and (T'_1, T'_2) have identical topology, but the topological transitions taking place throughout the geodesic from T_1 to T_2 are not the same as those taking place in the geodesic from T'_1 to T'_2 .

The length of the geodesic from T_1 to T_2 consists of three line segments connected via two topological transitions where, in the first line segment, the green edge is completely contracted from length 2 until it disappears, while the orange edge is contracted from length 6 to length 4. This line segment has length $\sqrt{2^2 + 2^2} = 2\sqrt{2}$. In the second line segment, the orange edge is further contracted from length 4 until it disappears, while the pink edge appears and grows from nothing to length 4 (the illustrated tree where the pink and orange edges both have length 2 is an intermediate tree along the line segment). The length of the second line segment is $\sqrt{4^2 + 4^2} = 4\sqrt{2}$. In the third and final line segment, the pink edge continues to grow from length 4 to 6, and a blue edge appears and grows to length 2. The length of the third line segment is thus $\sqrt{2^2 + 2^2} = 2\sqrt{2}$. The complete length of the geodesic is the sum of the lengths of the line segments, that is $8\sqrt{2}$.

The geodesic from T'_1 to T'_2 , on the other hand, consists of two line segments connected by a single topological transition. In the first line segment, the green and orange edges are both contracted from lengths 6 and 1, respectively, until they disappear. This line segment has

² Code freely available from: <http://vm1.cas.unc.edu/stat-or/workspace/miscellaneous/provan/treespace/>.

length $\sqrt{6^2 + 1^2}$. In the second line segment, the blue and pink edges appear and grow from nothing to lengths 6 and 1, respectively, giving a line segment of length $\sqrt{6^2 + 1^2}$. Again, the length of the geodesic is the sum of lengths of the line segments, which gives a total length of $2\sqrt{6^2 + 1^2}$.

In terms of the ambient Euclidean space, the geodesic connecting T'_1 and T'_2 is not a straight line, because tree-space does not fill out the whole ambient Euclidean space. A straight line path from T'_1 to T'_2 in $(\mathbb{R}^{15})^S$ would have to pass through the orthant formed from the partitions/axes $AR|BCD$ and $ACD|BR$, so some intermediate trees would have to simultaneously contain i) an edge that splits the labels B, C and D off from A and the root, as well as ii) an edge that splits the edges A, C and D off from B and the root. Clearly, the partitions $\{A, R\}$ and $\{B, R\}$ cannot both happen in the same tree (they are incompatible); hence that orthant does not exist in tree-space. This is how tree-space corners appear, which geodesics have to pass around.

In Fig. 6, branch geometry is represented by branch length for illustrative purposes. We remind the reader that in our experiments, branch geometry is represented by shape in the form of landmark points.

2.5 A hierarchical labeling algorithm

Ideally, we would search through the whole airway tree T , test all admissible configurations T_L of the 20 segment leaf labels and select the one that optimizes line 9 in Algorithm 1. However, for an airway tree with as few as 100 branches, the search space size is on the order of 100^{20} , which is too large to handle. In order to ensure computational feasibility, we choose a hierarchical subtree approach, where labels of different generations are added subsequently, as described in Fig. 9 and Algorithm 3. Here, more shallow branches are treated as leaves in the first steps of the algorithm, which works its way down to the segments. In each step of the hierarchical label placement, the optimal branches for the given set of labels is selected as those giving the minimal sum of distances to a set of training trees.

In each step of the hierarchical labeling, a specific set of descendants are assigned to a specific already labeled branch, as described in Algorithm 2. The descendants of each branch are determined by the hierarchy as in Fig. 9. For instance, in line 5 of Algorithm 3, the LMB branch has already been assigned, the subtree rooted at LMB is extracted, and the algorithm attempts to assign leaf labels $X = \{L6, LLB, LUL\}$ in any possible configuration spanning two generations below the LMB. The choice of searching 3 generations in some cases and 2 in others is a tradeoff between having enough space in the tree to assign all branches in a given step, and having a small enough tree to limit the number of possible configurations for the sake of computational speed. This tradeoff depends on the number of labels assigned in a step as well as the chance of having higher order nodes

in that particular subtree, as higher order nodes give more branches per generation.

The optimal configuration L of the label set X is selected based on geodesic distances to training trees. For each leaf-labeled tree T_i in the training set and each T_L , compute the geodesic distance $d(T_i, T_L)$ between the trees T_i and T_L in the tree-space defined in Section 2.4. A labeling of T with label set X is extracted by choosing the labeled tree T_{labeled} among the T_L that satisfies:

$$T_{\text{labeled}} = \arg \min_{T_L} \sum_{i \in I} d(T_i, T_L). \quad (2)$$

The step in line 6 of Algorithm 3 is different from the others. Since the RUL branch is not always present, it is problematic to root a hierarchy subtree at the RUL as is done in [28]. Thus, rather than searching the tree below the RUL branch, assigned in line 4, we search the tree obtained from the subtree rooted at the parent of RUL by removing the subtree rooted at BranchInt.

The choice of "leaves" used at the different steps in the hierarchy was made in order to minimize the number of "leaves" used while only using "leaves" that actually appear in as many people as possible. In a different application we would recommend a similar strategy.

3 EXPERIMENTAL METHODS

We evaluate three different aspects of labeling performance. First, we evaluate labeling accuracy, defined as the ability to assign labels to the same branches as clinical experts. Second, we evaluate labeling reproducibility, defined as the ability to assign labels to the same branches in airway trees extracted from repeated scans of the same subject. Third, we evaluate the dependence of accuracy and reproducibility on COPD diagnosis and severity. In all three aspects, the performance of the algorithm is compared to the average performance of clinical experts.

3.1 Data

The data used in the experiments comes from the Danish Lung Cancer Screening Trial [30]. It consists of low-dose (120 kV and 40 mAs) pulmonary CT scans and lung function measurements. The scans were obtained from a Multi Detector CT scanner (16 rows Philips Mx 8000), reconstructed using a hard algorithm (kernel D) with a resolution of approximately $0.78\text{mm} \times 0.78\text{mm} \times 1\text{mm}$. The lung function measurements, used to determine COPD severity, were performed using a computerized system (Spirotrac IV, Vitalograph) according to recommendations by the European Respiratory Society [31] without the use of bronchodilation.

The airway lumen surface was extracted from the images using the locally optimal path approach of [32], which was then refined using the optimal surface approach of [33]. Afterwards centerlines were computed by front propagation within the refined lumen surface

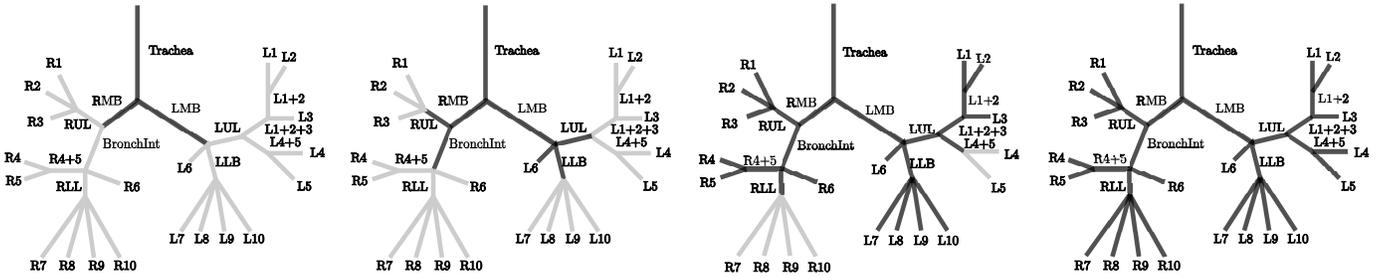


Fig. 9. Hierarchical labeling: In each step, search through a subtree for an optimal alignment of a subset of labels, obtaining a leaf-labeled subtree of the segmented airway tree similar to the trees shown in black. The real tree topology may differ; the figure only illustrates the stepwise hierarchy.

Algorithm 2 The labeling made in each hierarchical iteration.

- 1: **Input:** Root label r , leaf label set X , tree T with root label assigned, number of search generations g , training set $\{T_i | i \in I\}$ of fully labeled trees.
 - 2: # Generate leaf-labeled training trees:
 - 3: **for** $i \in I$ **do**
 - 4: $T_i \leftarrow$ subtree of T_i rooted at r
 - 5: $T_i \leftarrow$ subtree of T_i spanned by the labels X
 - 6: **end for**
 - 7: # Perform labeling:
 - 8: $\tilde{T} \leftarrow$ subtree of T rooted at r
 - 9: $\tilde{T} \leftarrow$ subtree of \tilde{T} of depth g
 - 10: Generate set \mathcal{L} of all possible admissible labelings of \tilde{T} with label set X , denoted $L \in \mathcal{L}$.
 - 11: **for** suggested labelings $L \in \mathcal{L}$ **do**
 - 12: $T_L =$ subtree of \tilde{T} spanned by the labels assigned by L
 - 13: Compute distances $\{d(T_L, T_i) | i \in I\}$
 - 14: Compute $f_L = \sum_{i \in I} d(T_L, T_i)$
 - 15: **end for**
 - 16: Labeled subtree $T_X = \operatorname{argmin}_{L \in \mathcal{L}} f_L$
 - 17: $T \leftarrow$ original tree T with labels X transferred from T_X .
 - 18: **Output:** T
-

Algorithm 3 For computational speed, the labeling is split into a set of hierarchical subtree labeling steps.

- 1: **Input:** Unlabeled tree \mathbf{T} .
 - 2: Label the first branch in the airway tree as the trachea.
 - 3: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{Trachea}$, $X = \{\text{LMB}, \text{RMB}\}$, $T = \mathbf{T}$, $g = 3$.
 - 4: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{RMB}$, $X = \{\text{RUL}, \text{BronchInt}\}$, $T = \mathbf{T}$, $g = 2$.
 - 5: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{LMB}$, $X = \{\text{L6}, \text{LLB}, \text{LUL}\}$, $T = \mathbf{T}$, $g = 2$.
 - 6: $\mathbf{T} \leftarrow$ tree returned by a modified³ version of Algorithm 2 with $X = \{\text{R1}, \text{R2}, \text{R3}\}$, $T = \mathbf{T}$, $g = 3$ or 2 .
 - 7: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{BronchInt}$, $X = \{\text{R4}, \text{R5}, \text{RLL}, \text{R6}\}$, $T = \mathbf{T}$, $g = 2$.
 - 8: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{LLB}$, $X = \{\text{L7}, \text{L8}, \text{L9}, \text{L10}\}$, $T = \mathbf{T}$, $g = 2$.
 - 9: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{LUL}$, $X = \{\text{L1}, \text{L2}, \text{L3}, \text{L4+5}\}$, $T = \mathbf{T}$, $g = 3$.
 - 10: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{RLL}$, $X = \{\text{R7}, \text{R8}, \text{R9}, \text{R10}\}$, $T = \mathbf{T}$, $g = 3$.
 - 11: $\mathbf{T} \leftarrow$ tree returned by Algorithm 2 with input $r = \text{L4+5}$, $X = \{\text{L4}, \text{L5}\}$, $T = \mathbf{T}$, $g = 2$.
 - 12: **Output:** Labeled tree \mathbf{T} .
-

as described in [34]. The resulting centerlines were disconnected at bifurcations and so Dijkstra's algorithm was used to connect them along shortest paths within an inverted distance transform of the refined lumen surface. The airway extraction was manually validated on 32 CT scans all from different subjects randomly chosen from the Danish Lung Cancer Screening Trial.

The software used was developed for the EXACT'09 airway challenge [34]. The average branch count was 188.6 ± 60.8 and tree length (excluding trachea and main bronchi) was 231.9 ± 81.0 cm, of which only $2.22 \pm 2.04\%$ was false. The method has previously been shown [32] to compare favorably with one of the best methods from

the EXACT'09 airway challenge [34].

A data set of 80 airway tree centerlines from 40 subjects scanned at two time-points with intervals of 5 years, were manually assigned segment labels $L1 - L10$ and $R1 - R10$ by two experts in pulmonary medicine (LHT and AD) and one in radiology (MMWW). The labels were assigned according to [35] and [36]. The experts were allowed to assign the same label to multiple branches in cases where they were unsure. The manual labeling was done using in-house developed software, simultaneously showing the segmented airway and centerline, which can be rotated, panned and zoomed, as well as a CT cross-section perpendicular to and centered on any given point of the airway. The remaining labels seen in Fig. 2 were deduced from the segment labels.

COPD severity was defined according to the GOLD standard [37], from the averaged lung function at both time-points. Out of the 40 subjects, there were 9 subjects with no airflow limitation, and 11 with mild, 11 with moderate, and 9 with severe COPD. We will denote the groups as GOLD 0, GOLD 1, GOLD 2, and GOLD 3, respectively.

The algorithm was further tested in a large longitudinal data set including all the subjects from the Danish Lung Cancer Screening Trial [30] who had at least two usable scans. For this data set lung CT image registration [38] was used to automatically determine reproducibility, as described in Section 4.3, and so it was important that the images could be registered well. A scan was therefore deemed non-usable if the lungs were not entirely contained within the image or if bowel air was erroneously included within lung segmentations (lung segmentation method and manual validation are described in [39]). This resulted in the inclusion of 1900 subjects of which, based on average lung function measurements over all time-points, 975 belonged to GOLD 0, 495 belonged to GOLD 1, 391 belonged to GOLD 2, and 38 belonged to GOLD 3. There was a single subject with very severe COPD (what would otherwise be GOLD 4), which was included in the GOLD 3 group. Each of these subjects had an average of 4.6 ± 0.7 usable scans approximately evenly distributed over a period of 5 years.

3.2 Implementation

The labeling algorithm was implemented in MATLAB, using tree distance computations implemented in Java⁴. For the annotated dataset, the airway trees had 181 branches on average, and the whole labeling took roughly 10 minutes per tree running on a laptop with a single 2.40 GHz processor using no more than 3 GB RAM per labeling.

3. In the modified version, rather than use the subtree rooted at *RUL*, we use the subtree obtained from the subtree rooted at the parent of *RUL* by removing the subtree rooted at *BronchInt*, as also discussed in the main text.

4. Code freely available from <http://vm1.cas.unc.edu/stat-or/workspace/miscellaneous/provan/treespace/>.

Label	Accuracy		Reproducibility		Labeled airways	
	Automatic	Expert	Automatic	Expert	Automatic	Expert
R1	89.2%	87.9%	95%	91.1%	80.0	80.0
R2	87.5%	84.2%	97.5%	90%	80.0	80.0
R3	87.6%	87.2%	97.5%	86.7%	80.0	80.0
R4	90.4%	88.7%	92.5%	86.4%	80.0	79.3
R5	86.9%	84.3%	90%	82.1%	80.0	79.0
R6	91.8%	93.3%	97.5%	91.5%	80.0	80.0
R7	77.7%	79.0%	84.6%	85.2%	76.0	79.7
R8	72.8%	75.8%	69.2%	79.2%	76.0	80.0
R9	63.2%	67.5%	53.8%	55.7%	76.0	80.0
R10	59.1%	64.3%	51.3%	55.8%	76.0	80.0
L1	64.0%	54.4%	75%	58.9%	79.9	79.7
L2	65.8%	60.1%	70%	62.2%	79.9	79.7
L3	66.0%	59.3%	75%	64.2%	79.9	80.0
L4	69.4%	78.3%	72.5%	80.7%	79.9	80.0
L5	73.2%	84.6%	70%	86.7%	79.9	80.0
L6	99.6%	99.2%	100%	99.2%	80.0	80.0
L7	62.8%	53.8%	82.5%	63.6%	80.0	77.3
L8	54.9%	48.6%	87.5%	57.1%	80.0	79.7
L9	53.4%	58.3%	72.5%	62.2%	80.0	80.0
L10	58.3%	57.1%	80%	62.5%	80.0	80.0
Trachea	100.0%	100.0%	100.0%	100.0%	80.0	80.0
LMB	100.0%	100%	100%	100%	80.0	80.0
LUL	100.0%	100%	97.5%	97.5%	80.0	80.0
LB1+2	62.9%	50%	70.6%	65.3%	61.8	57.0
LB4+5	92.3%	89.6%	95%	91.7%	79.9	78.7
LLB	99.2%	98.3%	100%	98.3%	80.0	79.3
RMB	100.0%	100%	100%	100%	80.0	80.0
RUL	97.4%	95%	100%	100%	78.0	76.0
BronchInt	99.9%	99.7%	100%	100%	80.0	80.0
RB4+5	95.8%	95.4%	95%	95%	80.0	78.0
RLL	93.3%	96.7%	95%	96.7%	80.0	79.7
LB1+2+3	92.0%	81.7%	94.6%	89.9%	74.9	69.0
Segmental average	73.7	73.3	80.8	75.0	79.2	79.7
Total average	$\pm 4.8\%$	$\pm 9.7\%$	$\pm 16.3\%$	$\pm 10.2\%$	79.2	79.7
Total average	81.5	80.4	86.4	82.4	78.7	78.5
average	$\pm 3.5\%$	$\pm 7.1\%$	$\pm 13.1\%$	$\pm 7.2\%$	78.7	78.5

TABLE 1

The mean accuracy of the algorithm was computed from 10 repetitions of 10-fold cross validation, and the mean accuracy of an expert was averaged over all three pairs of experts. The third and fourth columns contain the mean reproducibility of the algorithm and an expert, respectively. The mean number of airways in which a given label was assigned, was averaged over 10 cross-validation runs or three experts, respectively.

Labeling experiments on the annotated dataset were performed using 10-fold cross validation, where both scans of any individual were always contained in the same fold. Thus, for each test set fold of 8 airway trees from 4 patients, the training set was made up of 72 airway trees from 36 patients. Each tree was labeled separately by the three medical experts, but sometimes the medical experts would, when in doubt, place the same label on two different branches. In these cases two leaf-labeled training trees would be generated, one for each option. In other cases, some labels were not assigned by the expert, in which case the corresponding training subtree would not be generated. This resulted in 231-280 training airway trees from the 80 scans, with different numbers at different steps of the hierarchy.

3.3 Evaluation criteria

3.3.1 Labeling accuracy

The accuracy of the automatic labeling, defined as its ability to agree with a clinical expert, was assessed using 10 labeling runs of 10-fold cross validation with randomized folds. For each airway, the average success

rate was computed out of the number of labels assigned by either algorithm or expert (meaning that if neither the algorithm nor the expert assign a given label, then this label does not contribute to the success rate of that airway tree at all).

It is not obvious how labeling accuracy should be assessed. In some cases where experts were not certain, or judged that an anatomical branch had been split into two branches by the segmentation algorithm, they would assign the same segment label to multiple branches (the three experts did this in 34, 12 and 5 of the 80 airway trees, respectively). In other cases branches were missing, either anatomically or from the segmentation, so that some labels were not assigned (the three experts did this in 10, 1 and 4 of the 80 airway trees, respectively). Missing label assignments happened both in expert and automatic labelings. Sometimes branches would be missing in the airway tree, making label assignments impossible. Other times, the topology of the airway made certain non-segment labels impossible. Thus, Table 1 contains average counts for how many times labels were assigned by the algorithm and the experts, respectively.

In order to fairly assess all cases, we gave, for the i^{th} airway tree and each label $x \in X = \{\text{L1}, \dots, \text{L10}, \text{R1}, \dots, \text{R10}\}$, the assignment by method M_1 a correctness percentage $s_i(x, M_1, M_2)$ with respect to method M_2 . M_1 could be either an expert or the automatic labeling, and M_2 was always an expert. The correctness percentage was defined as follows: In the i^{th} tree T_i , let $M_1(T_i, x)$ denote the set of branches assigned label x by method M_1 and $M_2(T_i, x)$ the set of branches assigned label x by method M_2 . Define the correctness $s_i(x, M_1, M_2)$ of label x using method M_1 with respect to method M_2 in the i^{th} airway tree as:

$$s_i(x, M_1, M_2) = 100 \cdot \frac{2 \cdot |M_1(T_i, x) \cap M_2(T_i, x)|}{|M_1(T_i, x)| + |M_2(T_i, x)|} \%,$$

We assume cases where the label was not assigned by either method, that is $|M_1(T_i, x)| = |M_2(T_i, x)| = 0$, to be due to missing branches and thus leave them out of the total summary shown in Table 1.

3.3.2 Labeling reproducibility

In order to test scan-rescan reproducibility of the expert and automatically assigned labels, the two CT scans of each subject were registered using deformable image registration as described in [38], and the labeled airway branches were manually investigated for possible matches in the resulting common coordinate system. Let T_i^1 and T_i^2 be two trees corresponding to the i^{th} subject's airway at time-points 1 and 2, and let $M(T_i^1, x) \subseteq E_i^1$ and $M(T_i^2, x) \subseteq E_i^2$ be sets of branches assigned label x by the method M in T_i^1 and T_i^2 , respectively. Denote by $R(M(T_i^1, x), M(T_i^2, x))$ the matched subset of these branches.

We define the reproducibility of label x using the

method M within the trees T_i^1 and T_i^2 of subject i

$$r_i(x, M, T_i^1, T_i^2) = 100 \cdot \frac{2 \cdot |R(M(T_i^1, x), M(T_i^2, x))|}{|M(T_i^1, x)| + |M(T_i^2, x)|} \%.$$

To avoid evaluating effects of missing branches due to segmentation problems, cases where the label was not assigned in either time-point, that is $|M(T_i^1, x)| = |M(T_i^2, x)| = 0$ are left out of the total summary. The same holds for cases where only one time-point was labeled with x , which without loss of generality can be assumed to be T_i^1 , but only if matching branches did not exist in the other time-point, that is $R(M(T_i^1, x), M(T_i^2, x)) = \emptyset$.

3.3.3 Labeling reproducibility on large longitudinal dataset

Reproducibility of the automatic approach on the larger data set was tested by labeling all 8724 scans using the manually labeled airway trees as a training set, in which the results for the 40 subjects that had manual annotations were computed in a leave-one-subject-out fashion. All scans of each subject were registered [38] and branches appearing in multiple images were matched, in a similar fashion to what was described in the previous section. However, rather than manually detecting matched branches, which would be very time consuming for a data set of this size (8724 trees), matching was done automatically. The details of this automatic matching approach have previously been published in [40].

It is not obvious how labeling reproducibility should be defined in a subject with more than two time-points. As an example, consider a case where three out of five time-points agree on one assignment of the label x , and the remaining two time-points agree on another. Taking such cases into account, we define reproducibility of assigning the label x in terms of percentage agreement with the majority labeling (if two labelings are both majority, one of them is just selected).

To define reproducibility analytically, let T_i^1, \dots, T_i^n denote the airway trees of subject i at n different time-points, and let $M(T_i^1, x), \dots, M(T_i^n, x)$ be the branches assigned label x by method M in each time-point. In addition let the set of time-points where the assignment of the label x matches the assignment in time-point j be:

$$N_i(M, x, j) = \left\{ k \in \{1, \dots, n\} \mid R(M(T_i^j, x), M(T_i^k, x)) \neq \emptyset \right\}.$$

A time point where the majority labeling occurs is then given by:

$$j_{\max}(x, M) = \operatorname{argmax}_{j \in \{1, \dots, n\}} |N_i(x, j, M)|.$$

To avoid evaluating effects of missing branches due to segmentation problems, time-points $k \in \{1, \dots, n\}$ are left out if x has not been assigned, that is $M(T_i^k, x) = \emptyset$, and none of the other branches match the branch labeled with x in a majority labeled time-point, that is

$R(M(T_i^{j_{\max}(x,M)}), x), T_i^k) = \emptyset$. The remaining time-points are denoted by $P_i(x, M)$:

$$P_i(x, M) = \left\{ k \in \{1, \dots, n\} \mid R(M(T_i^{j_{\max}(x,M)}), x), T_i^k) \neq \emptyset \text{ or } M(T_i^k, x) \neq \emptyset \right\}.$$

We then define reproducibility of a label x in subject i by method M as the percentage of time-points agreeing with the majority labeling out of the total amount of included time-points:

$$r_i(x, M) = 100 \cdot \frac{|N_i(x, j_{\max}(x, M), M)|}{|P_i(x, M)|} \%,$$

where $|N_i(M, x, j_{\max}(x, M))| > 1$, otherwise $r_i(x, M) = 0$. Cases with less than two included time-points, that is $|P_i(x, M)| < 2$, are left out. Note that in the case of two time-points, this definition of reproducibility is the same as the one defined in Section 4.2.

4 RESULTS

Fig. 10 shows two labeling results visualized; examples of both rare and more common topology. In the case with rare topology, the labels L1, L2, L3, L10 and L1+2 did not overlap with expert labels; R1 and R3 overlapped with one expert; R7, R9, R10, L7, L8, and L9 overlapped with two experts; and the remaining 19 labels had perfect overlap. In the case with more common topology, the labels L7, L8, and L10 overlapped with one expert; L9 and RLL overlapped with two experts; and the remaining 27 labels had perfect overlap.

4.1 Labeling accuracy

Table 1 shows a summary of the results. On average, the automatic labeling agreement with an expert was $73.7 \pm 4.8\%$ on the segment branches, and $81.5 \pm 3.5\%$ overall. This is not significantly different from the average expert agreement with an expert, which was $73.3 \pm 9.7\%$ on the segment labels, and $80.4 \pm 7.1\%$ overall ($p = 0.94$ and $p = 0.77$ in Mann-Whitney U-tests).

Fig. 11 shows labeling accuracy stratified by COPD severity. Spearman's correlation test shows no significant correlation between the average agreement with an expert and the presence and severity of COPD ($\rho = -0.18$, $p = 0.11$ on all labels, $\rho = -0.20$, $p = 0.069$ on segment labels). Similarly, there is no correlation between the average agreement between experts, and presence and severity of COPD ($\rho = -0.12$, $p = 0.45$ on all labels, $\rho = -0.085$, $p = 0.60$ on segment labels).

4.2 Labeling reproducibility

Table 1 shows a summary of the results.

On average, the reproducibility of the automatic labeling was $80.8 \pm 16.3\%$ on the segment labels, and $86.4 \pm 13.1\%$ overall, which is significantly better than the reproducibility of the experts, which was $75.1 \pm 14.8\%$ on

the segment labels, and $82.4 \pm 10.6\%$ overall ($p = 0.021$ and $p = 0.022$ in Mann-Whitney U-tests).

Fig. 11 shows labeling reproducibility stratified by COPD severity. Spearman's correlation test shows significant correlation between reproducibility of the automatic approach and severity of COPD ($\rho = -0.34$, $p = 0.031$ on all labels; $\rho = -0.36$, $p = 0.024$ on segment labels). Spearman's correlation test shows, however, no significant correlation between the average reproducibility of the expert labeling and the severity of COPD ($\rho = -0.085$, $p = 0.604$ on all labels; $\rho = -0.049$, $p = 0.764$ on segment labels).

4.3 Reproducibility on large longitudinal data set

Fig. 12 shows a summary of the results.

On average, the reproducibility of the automatic labeling on the large longitudinal data set was $82.5 \pm 12.0\%$ on the segment labels, and $86.9 \pm 9.8\%$ overall.

There was a significant correlation between the reproducibility and severity of COPD ($\rho = -0.158$, $p < 0.001$ on all labels; $\rho = -0.163$, $p < 0.001$ on segment labels). Airway segmentations can be less complete in more diseased subjects [4], [12], and we also observed a significant correlation between number of extracted branches and severity of COPD ($\rho = -0.444$, $p < 0.001$) and between number of extracted branches and reproducibility ($\rho = 0.287$, $p < 0.001$ on all labels; $\rho = 0.308$, $p < 0.001$ on segment labels). Fig. 13 shows reproducibility plotted against number of segmented branches. To investigate whether the algorithm was truly sensitive to disease and not just missing branches, we generated a normalized reproducibility by subtracting the mean predicted reproducibility, predicted from the number of segmented branches using a locally weighted mean (Loess Curve) also shown in Fig. 13, from the actual reproducibility. This normalized reproducibility did not significantly correlate with disease ($\rho = -0.043$, $p = 0.059$ on all labels; $\rho = -0.032$, $p = 0.166$ on segment labels).

5 DISCUSSION

We have presented a novel atlas-based algorithm for assigning anatomical branch labels in airway trees extracted from CT. Through detailed experimental validation we show that the performance of the algorithm is as good as the performance of the clinical experts. In particular, the accuracy and reproducibility of the algorithm is over 90% on the non-segment branches as well as on the L6 and R6 branches, and for many of the branches even over 95%. The labeling is fast, uses little memory, and easily runs on a standard laptop.

We have chosen to use the geodesic tree-space distance between pairs of leaf-labeled trees. In principle, any other distance between leaf-labeled trees could have been used in its place, e.g. the weighted Robinson-Foulds metric [41] which is a related distance measure in the same tree-space. The geodesic distance is attractive because, in addition to the fact that polynomial time

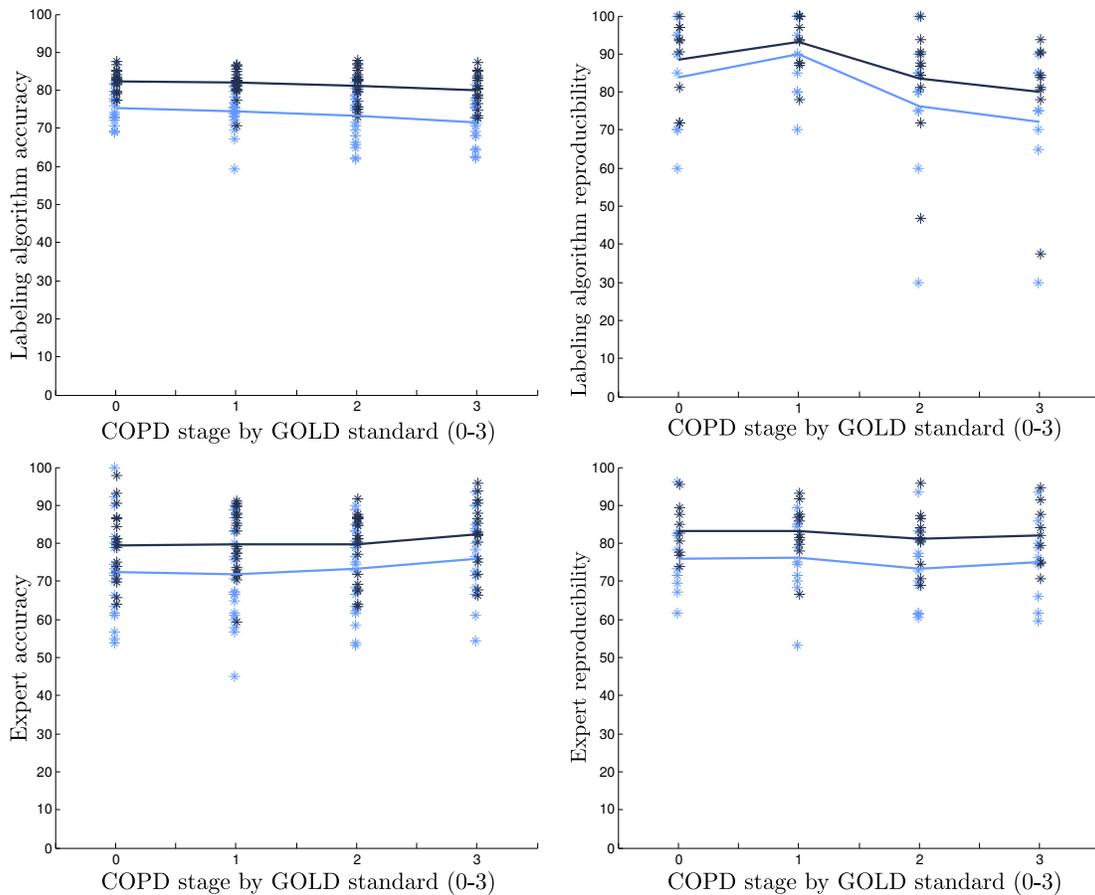


Fig. 11. Average labeling accuracy (left) and reproducibility (right) for segment labels (light blue) and all labels (dark blue), stratified by COPD severity for the algorithm (top) and experts (bottom). Each * corresponds to accuracy/reproducibility for one subject. The lines interpolate mean accuracies and reproducibilities for each GOLD group.

hierarchical labeling scheme, particularly one informed by an analysis of seen topologies, or as discussed below, of the airway subtrees where the experts performed better. In cases with less severe disease, it is straightforward to add additional measurements such as lumen diameter into the distance function, which could lead to increased performance.

As seen in Fig. 10 the method can deal with previously unseen topology. Besides cases with rare anatomy, this property is also relevant due to the large natural anatomical variation and the cases of segmentation or centerline errors, such as spurious branches. In our expert-labeled dataset, the leaf-labeled trees with exactly 20 segment branches extracted from the three experts' manually labeled trees give rise to 137 different topologies as illustrated in Fig. 14. Of these topologies, 73 only appear a single time, and a few are shared by a good number of airway trees (the 11 most common topologies are shared by a total of 84 trees). The large number of observed topologies is likely due to the many possible combinations of different topologies found in the different lobes; we expect that many of these globally different topologies share large parts of their topolog-

ical structure. This topological variation is one of the challenges that the tree-shape model is particularly well suited to handle, as it allows distances to be computed between topologically different trees. In particular, this allows us to use the entire training set when labeling an unseen tree, in spite of the unseen tree being topologically different from many or even all of the training trees. This advantage is illustrated by our strong performance in spite of the topological diversity of our dataset.

It is noteworthy that the experts and the algorithm perform well in different parts of the airway tree. In particular, the algorithm is far more reproducible than the experts in the left upper and lower lobes (L1-L3 and L7-L10) (70%-87.5% versus 57.1%-64.2%). These branches are also the hardest to label according to expert accuracy (< 60.1%). It is possible that biological variation of shape and topology confuses the experts, making their labels more random, which would lead to both low accuracy and reproducibility. The algorithm might either be more tolerant of biological variation or more consistent in the types of errors made. On the other hand, the experts perform better than the algorithm in the left middle and right lower lobes (L4-L5 and R7-R10). These branches

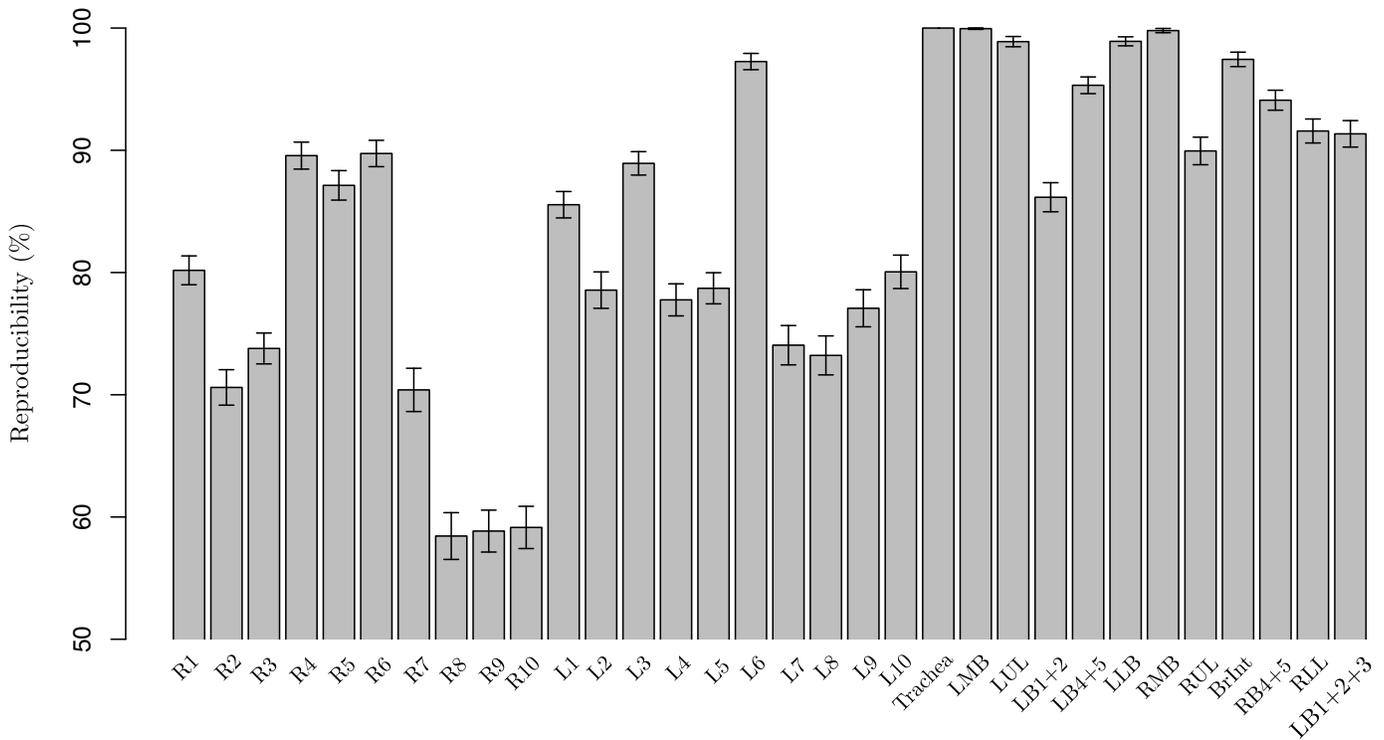


Fig. 12. Reproducibility of the automatically assigned individual labels in the large longitudinal data set. Error bars indicate 95% confidence intervals.

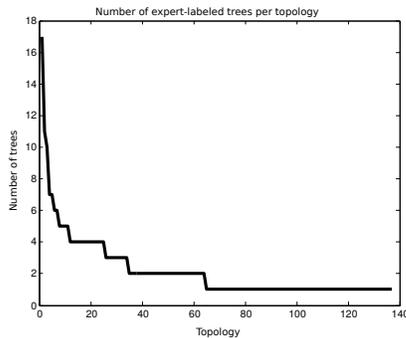


Fig. 14. Number of airway trees per leaf-labeled topology with 20 segmental leaf labels extracted from our expert-labeled dataset.

belong to the subtrees of LB4+5 and RLL, which are the least accurately found branches of the subset of branches used as steps in the hierarchical approach. Some amount of error is therefore probably due to the hierarchical search strategy. It is possible that better results could be obtained by for instance searching for L4-L5 and R7-R10 label configurations within subtrees depending on more than one choice of the LB4+5 and RLL branches.

It is interesting to compare the estimated reproducibility in the small and large data sets. In general the trends are the same. For instance in both data sets R6 and L6 are among the most reproducibly assigned segment labels and lower lobe segment labels are in general less reproducibly assigned than upper and middle lobe segment

labels. The mean reproducibilities of the two data sets are also almost identical. It should be noted, however, that the automatic matching method [40] requires an accurate registration and unlike with the manual matching the overall topology and shape of the tree is not taken into account. Because of this the true reproducibility is probably underestimated in the larger data set. However, the larger data set also has relatively fewer severe COPD cases, which should mean the reproducibility is higher because of more completely segmented airway trees. Despite this, the similarity of the results indicate that the performance of the algorithm generalizes to new (albeit similar) data.

5.2 Labeling performance and COPD stage

Our experiments on the annotated dataset show that labeling reproducibility decreases significantly with increased COPD severity, while labeling accuracy does not. The difference in results may be caused by the mathematical definition of accuracy and reproducibility.

Labeling accuracy measures the ability of a method to agree with a human observer on the same segmented airway tree. If the underlying labeling “algorithm” used by the method and the human observer are identical, accuracy will be perfect and there will be no dependence on COPD severity even if both are physiologically incorrect. Reproducibility, on the other hand, measures the ability of *one* method to identically label segmented airway trees from two different scans of the same subject. If one of the segmentations is missing branches that

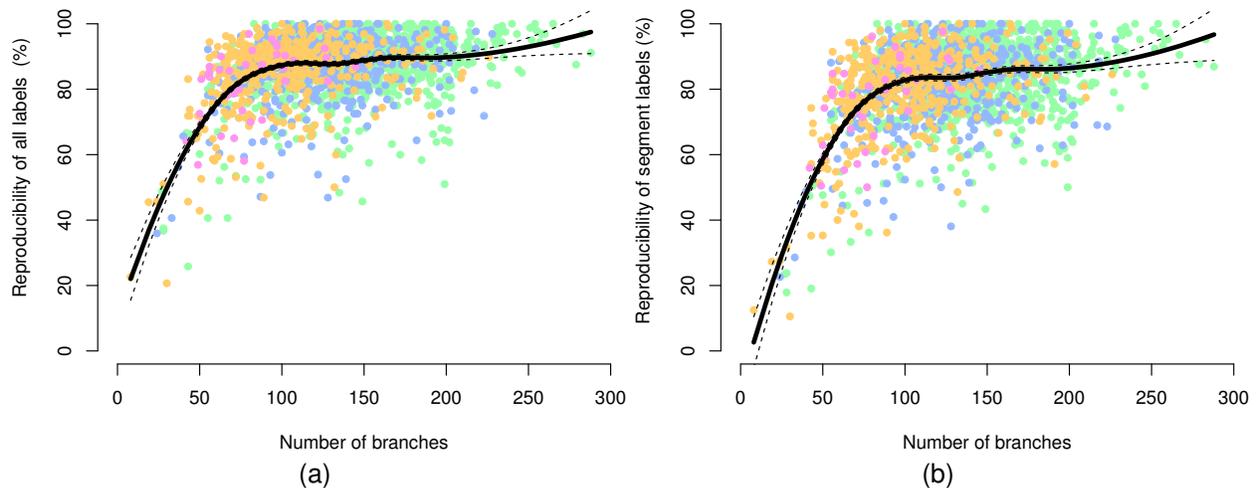


Fig. 13. Reproducibility of all labels left and segment labels right as a function of the number of segmented branches. Colors indicate GOLD group, with GOLD 0 in green, GOLD 1 in blue, GOLD 2 in yellow, and GOLD 3 in pink. The black line is a locally weighted mean (Loess Curve) with 95% confidence intervals.

are labeled by the other, then reproducibility cannot possibly be perfect. Thus, since the number of segmented branches depends on COPD level, it is expected that reproducibility depends on COPD level as well, while this is not necessarily true for accuracy.

This is supported by our reproducibility experiments on the large longitudinal data set. Here, reproducibility does again depend on COPD level. However, there is also a very clear correlation between reproducibility and the number of branches detected by the segmentation algorithm. When the effect of branch number is taken into account by subtracting the expected reproducibility based on number of branches, the correlation between reproducibility and COPD level disappears. We conclude that the correlation between COPD level and reproducibility is not an artifact of the labeling algorithm directly, but a result of segmentation problems. This is very natural when many branches close to the true named branches are missing, and it has several consequences: if the named branch is missing, then any attempt to assign the corresponding label will fail, and if one but not both children of a named branch is missing, then the branch will appear longer in the segmentation than it should, making it harder to assign labels based on branch features such as shape.

Of course, one solution to the dependence on segmentation quality and, indirectly, disease, is to develop better segmentation algorithms. A more pragmatic approach, however, could be to introduce label probabilities based on geodesic airway tree distances, giving an option of assigning fewer labels when higher accuracy is needed, in a similar way as done in [11]. This could also be used to decrease the false positive rate on difficult branches.

5.3 Applicability to other segmentation algorithms and data

Our experiments are performed on airway segmentations of high quality. A relevant question is whether the algorithm generalizes to other scans and segmentation algorithms, both in its current trained form and retrained on new data. We expect the trained algorithm to be robust to common differences in the output of segmentation algorithms such as differences in surface shape, as the only branch shape feature used is 5 landmark points sampled along the centerline shape. The robustness of the algorithm in the presence of disease is a positive indicator for robustness to differences in branch shape segmentation.

Potential errors in alternative segmentations include missing and spurious branches. In Sec. 5.2 we have found that a somewhat decreased labeling performance in COPD can be explained by fewer detected branches in subjects with COPD. This indicates that the algorithm is somewhat sensitive to missing segmental branches. On the other hand, we would expect it to be relatively robust to spurious branches. Spurious branches are removed during the process generating the subtree spanned by the labels, which means there is limited effect on the training trees. There might be an effect on the test tree, but this effect is significantly reduced by both the subtree spanned by the labels and the backpropagation of labels to the most shallow possible parent.

5.4 Relation to alternative methods

We note that higher labeling accuracy percentages than ours are reported in the literature, 97.1%, 90% and 83% on all branch labels in [17, high dose CT], [9] and [11]; and 77% on segment labels [11]. There are several reasons for this. First, as noted above, we specifically aim to evaluate our performance on patients with disease,

and our dataset consists of 77.5% COPD patients, while experiments found in the literature include much fewer, if any, subjects with an airway disease. A lower performance on our dataset is thus expected, since our experiments prove that labeling performance significantly depends on disease when segmentation problems are not accounted for.

Second, we aim to evaluate our ability to assign 20 segment labels. In comparison, both [11] and [9] use fewer than 20 segment labels (16 and 19, respectively) and more intermediate (easier) labels (13 as opposed to our 12), which presumably gives higher overall performance summaries. We note in particular that [11] and [9] leave out the segment label sets $\{L1-L2, L7-L8\}$ and $\{L7\}$, respectively, which are also found challenging in our experiments.

Third, we wanted our evaluation to realistically reflect how well we can expect to perform on data which has never been labeled by an expert. To achieve this, we did not reject any assigned labelings, as opposed to [17] and [11], which aim to avoid performing uncertain labelings. In particular, the 97.1% success rate [17] is among branches that have been labeled identically by three experts, which means that difficult labeling problems are weeded out of the experiment. Such an evaluation is only possible if the airway tree has already been labeled by three experts, and similar accuracies can naturally not be expected on unseen data, such as in clinical applications. In [11], an estimate is made of the probability of the label assignment, and here, a threshold can be applied to choose not to assign labels when certainty is low.

On average (including erroneous labelings), we assign 98.4% of 32 used labels, whereas [17], [9] and [11] assign only 71%, 93%, and 83% of the 29, 32, and 32 used labels, respectively. For segment labels specifically, we assign 94.9% of 20 used segment labels, whereas [11] assigns 77% of the 16 segment labels used (our numbers are averaged over the 10 cross-validation runs). This variation in experimental setup makes it impossible to compare performance in a fair manner, because results on unassigned labels cannot be taken into account.

What we *can* conclude is that the proposed algorithm performs as well as or better than medical experts in terms of labeling accuracy and reproducibility. This is the best result we could possibly have hoped for given that our method is trained on labeling performed by medical experts. These conclusions are confirmed by our large-scale evaluation of reproducibility. Moreover, we quantify the dependence on performance on COPD level and show that any negative correlation between labeling performance and disease can be explained by segmentation error. To the best of our knowledge, no previous work has tested neither reproducibility nor dependence on disease, nor performed large-scale evaluations.

6 CONCLUSION

We present a new atlas-based algorithm for anatomical branch labeling of airway trees, based on geodesic tree-

space distances between airway trees. Using the distances, the algorithm evaluates how well a suggested branch labeling fits with a training set of labeled airway trees, and chooses the optimal labeling. The labeling performance is robust in patients with COPD, and is comparable in performance to that of experts in pulmonary medicine and radiology. As the algorithm only uses branch centerlines and tree topology, we expect it to generalize to other data sets consisting of similarly complete segmentations of human adults. Its reproducibility and robustness in patients with COPD emphasizes its suitability for use in clinical studies of localized CT-based airway measurements.

REFERENCES

- [1] M. Hasegawa, Y. Nasuhara, Y. Onodera, H. Makita, K. Nagai, S. Fuke, Y. Ito, T. Betsuyaku, and M. Nishimura, "Airflow limitation and airway dimensions in chronic obstructive pulmonary disease," *American Journal of Respiratory and Critical Care Medicine*, vol. 173, no. 12, pp. 1309–1315, 2006.
- [2] J. Petersen, M. Wille, L. Raklt, A. Feragen, J. Pedersen, M. Nielsen, A. Dirksen, and M. de Bruijne, "Effect of inspiration on airway dimensions measured in maximal inspiration ct images of subjects without airflow limitation," *European Radiology*, vol. 24, no. 9, pp. 2319–2325, 2014.
- [3] M. Lederlin, F. Laurent, Y. Portron, A. Ozier, H. Cochet, P. Berger, and M. Montaudon, "CT attenuation of the bronchial wall in patients with asthma: comparison with geometric parameters and correlation with function and histologic characteristics," *American Journal of Roentgenology*, vol. 199, no. 6, pp. 1226–33, 2012.
- [4] A. Diaz, C. Valim, T. Yamashiro, R. Estpar, J. Ross, S. Matsuoka, B. Bartholmai, H. Hatabu, E. Silverman, and G. Washko, "Airway count and emphysema assessed by chest CT imaging predicts clinical outcome in smokers," *CHEST Journal*, vol. 138, no. 4, pp. 880–887, 2010.
- [5] K. Mori, J. Hasegawa, Y. Suenaga, J. Toriwaki, H. Anno, and K. Katada, "Automated labeling of bronchial branches in virtual bronchoscopy system," in *MICCAI*, 1998, pp. 870–878.
- [6] K. Mori, J. Hasegawa, Y. Suenaga, and J. Toriwaki, "Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system," *IEEE Trans. Med. Imaging*, vol. 19, no. 2, pp. 103–114, 2000.
- [7] K. Mori, S. Ema, T. Kitasaka, Y. Mekada, I. Ide, H. Murase, Y. Suenaga, H. Takabatake, M. Mori, and H. Natori, "Automated nomenclature of bronchial branches extracted from ct images and its application to biopsy path planning in virtual bronchoscopy," in *MICCAI (2)*, 2005, pp. 854–861.
- [8] S. Gu, Z. Wang, J. M. Siegfried, D. Wilson, W. L. Bigbee, and J. Pu, "Automated lobe-based airway labeling," *Journal of Biomedical Imaging*, vol. 2012, Jan. 2012.
- [9] B. van Ginneken, W. Baggerman, and E. van Rikxoort, "Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2008*, ser. Lecture Notes in Computer Science, 2008, pp. 219–226.
- [10] K. Mori, S. Ota, D. Deguchi, T. Kitasaka, Y. Suenaga, S. Iwano, Y. Hasegawa, H. Takabatake, M. Mori, and H. Natori, "Automated anatomical labeling of bronchial branches extracted from CT datasets based on machine learning and combination optimization and its application to bronchoscope guidance," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009*, ser. Lecture Notes in Computer Science, 2009, vol. 5762, pp. 707–714.
- [11] P. Lo, E. van Rikxoort, J. Goldin, F. Abtin, M. de Bruijne, and M. Brown, "A bottom-up approach for labeling of human airway trees," in *Proceedings of MICCAI International Workshop on Pulmonary Image Analysis*, 2011.
- [12] J. Pu, J. K. Leader, X. Meng, B. Whiting, D. Wilson, F. C. Sciruba, J. J. Reilly, W. L. Bigbee, J. Siegfried, and D. Gur, "Three-dimensional airway tree architecture and pulmonary function," *Academic Radiology*, vol. 19, no. 11, pp. 1395–401, 2012.

- [13] M. O. Wielputz, M. Eichinger, O. Weinheimer, S. Ley, M. A. Mall, M. Wiebel, A. Bischoff, H. U. Kauczor, C. P. Heußel, and M. Puderbach, "Automatic airway analysis on multidetector computed tomography in cystic fibrosis: correlation with pulmonary function testing," *Journal of Thoracic Imaging*, vol. 28, no. 2, pp. 104–113, 2013.
- [14] W. K. Moon, J. G. Im, K. M. Yeon, and M. C. Han, "Tuberculosis of the central airways: CT findings of active and fibrotic disease," *American Journal of Roentgenology*, vol. 169, no. 3, pp. 649–53, 1997.
- [15] C. Pisupati, L. Wolff, W. Mitzner, and E. Zerhouni, "Tracking 3D pulmonary tree structures," in *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, Jun 1996, pp. 160–169.
- [16] M. W. Graham and W. E. Higgins, "Optimal graph-theoretic approach to 3D anatomical tree matching," in *International Symposium on Biomedical Imaging (ISBI)*, 2006, pp. 109–112.
- [17] J. Tschirren, G. McLennan, K. Palágyi, E. A. Hoffman, and M. Sonka, "Matching and anatomical labeling of human airway tree," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1540–1547, 2005.
- [18] H. Kitaoka, Y. Park, J. Tschirren, J. M. Reinhardt, M. Sonka, G. McLennan, and E. A. Hoffman, "Automated nomenclature labeling of the bronchial tree in 3D-CT lung images," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2002*, ser. Lecture Notes in Computer Science, 2002, pp. 1–11.
- [19] J. H. Metzger, T. Kröger, A. Schenk, S. Zidowitz, H.-O. Peitgen, and X. Jiang, "Matching of anatomical tree structures for registration of medical images," *Image and Vision Computing*, vol. 27, pp. 923–933, 2009.
- [20] H. Bogunovic, J. Pozo, R. Cardenas, L. San Roman, and A. Frangi, "Anatomical labeling of the Circle of Willis using maximum a posteriori probability estimation," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1587–99, 2013.
- [21] D. Smeets, P. Bruyninckx, J. Keustermans, D. Vandermeulen, and P. Suetens, "Robust matching of 3D lung vessel trees," in *Proceedings of MICCAI International Workshop on Pulmonary Image Analysis*, 2010, pp. 61–70.
- [22] T. Bülow, C. Lorenz, R. Wiemker, and J. Honko, "Point based methods for automatic bronchial tree matching and labeling," in *SPIE Medical Imaging*, vol. 6143, 2006, pp. 225–234.
- [23] J. Kaftan, A. Kiraly, D. Naidich, and C. Novak, "A novel multi-purpose tree and path matching algorithm with application to airway trees," in *SPIE Medical Imaging*, vol. 6143 I, 2006.
- [24] J. C. Ross, A. A. Diaz, Y. Okajima, D. Wassermann, G. R. Washko, J. Dy, and R. S. J. Estepar, "Airway labeling using a hidden markov tree model," in *International Symposium on Biomedical Imaging (ISBI)*, 2014, pp. 554–558.
- [25] A. Feragen, P. Lo, V. Gorbunova, M. Nielsen, A. Dirksen, J. Reinhardt, F. Lauze, and M. de Bruijne, "An airway tree-shape model for geodesic airway branch labeling," in *MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, 2011.
- [26] L. J. Billera, S. P. Holmes, and K. Vogtmann, "Geometry of the space of phylogenetic trees," *Advances in Applied Mathematics*, vol. 27, no. 4, pp. 733–767, 2001.
- [27] M. Owen and J. S. Provan, "A fast algorithm for computing geodesic distances in tree space," *ACM/IEEE Transactions on Computational Biology and Bioinformatics*, vol. 8, pp. 2–13, 2011.
- [28] A. Feragen, J. Petersen, M. Owen, P. Lo, L. H. Thomsen, M. M. W. Wille, A. Dirksen, and M. de Bruijne, "A hierarchical scheme for geodesic anatomical labeling of airway trees," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2012*, ser. Lecture Notes in Computer Science, 2012, pp. 147–155.
- [29] P. Lo, J. Sparring, H. Ashraf, J. Pedersen, and M. de Bruijne, "Vessel-guided airway tree segmentation: A voxel classification approach," *Medical Image Analysis*, vol. 14, no. 4, pp. 527–538, 2010.
- [30] J. Pedersen, H. Ashraf, A. Dirksen, K. Bach, H. Hansen, P. Toennesen, H. Thorsen, J. Brodersen, B. Skov, M. Døssing, J. Mortensen, K. Richter, P. Clementsen, and N. Seersholm, "The Danish randomized lung cancer CT screening trial - overall design and results of the prevalence round," *Journal of Thoracic Oncology*, vol. 4, no. 5, pp. 608–614, May 2009.
- [31] M. R. Miller, R. Crapo, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, P. Enright, C. P. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, J. Wanger, and A. T. Force, "General considerations for lung function testing," *European Respiratory Journal*, vol. 26, no. 1, pp. 153–61, 2005.
- [32] P. Lo, J. Sparring, J. Pedersen, and M. Bruijne, "Airway tree extraction with locally optimal paths," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009*, ser. Lecture Notes in Computer Science, 2009, vol. 5762, pp. 51–58.
- [33] J. Petersen, M. Nielsen, P. Lo, L. H. Nordenmark, J. H. Pedersen, M. M. W. Wille, A. Dirksen, and M. de Bruijne, "Optimal surface segmentation using flow lines to quantify airway abnormalities in chronic obstructive pulmonary disease," *Medical Image Analysis*, 2014 (In Press). [Online]. Available: <http://dx.doi.org/10.1016/j.media.2014.02.004>
- [34] P. Lo, B. van Ginneken, J. M. Reinhardt, T. Yavarna, P. A. de Jong, B. Irving, C. Fetita, M. Ortner, R. Pinho, J. Sijbers, M. Feuerstein, A. Fabijanska, C. Bauer, R. Beichel, C. S. Mendoza, R. Wiemker, J. Lee, A. P. Reeves, S. Born, O. Weinheimer, E. M. van Rikxoort, J. Tschirren, K. Mori, B. Odry, D. P. Naidich, I. Hartmann, E. A. Hoffman, M. Prokop, J. H. Pedersen, and M. de Bruijne, "Extraction of airways from CT (EXACT'09)," *IEEE Transactions on Medical Imaging*, vol. 31, no. 11, pp. 2093–2107, 2012.
- [35] F. Netter, *Atlas of human anatomy*. Novartis: East Hanover: NJ, 1989.
- [36] H. Feneis, *Anatomisk billedordbog*. Forlaget Munksgaard, Copenhagen, 1995.
- [37] K. F. Rabe, S. Hurd, A. Anzueto, P. J. Barnes, S. A. Buist, P. Calverley, Y. Fukuchi, C. Jenkins, R. Rodriguez-Roisin, C. van Weel, and J. Zielinski, "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary," *American Journal of Respiratory and Critical Care Medicine*, vol. 176, no. 6, pp. 532–555, 2007.
- [38] V. Gorbunova, J. Sparring, M. Nielsen, A. Dirksen, P. Lo, and M. de Bruijne, "Mass preserving image registration for lung CT," *Medical Image Analysis*, vol. 16, no. 4, pp. 786–95, 2012.
- [39] H. Ashraf, P. Lo, S. B. Shaker, M. de Bruijne, A. Dirksen, P. Tønnesen, M. Dahlbäck, and J. H. Pedersen, "Short-term effect of changes in smoking behaviour on emphysema quantification by CT," *Thorax*, vol. 66, no. 1, pp. 55–60, 2011.
- [40] J. Petersen, V. Gorbunova, M. Nielsen, A. Dirksen, P. Lo, and M. de Bruijne, "Longitudinal analysis of airways using registration," in *Fourth International Workshop on Pulmonary Image Analysis*, 2011.
- [41] D. Robinson and L. Foulds, "Comparison of weighted labeled trees," in *Combinatorial Mathematics VI*, ser. Lecture Notes in Mathematics, 1979, vol. 748, pp. 119–126.
- [42] A. Feragen, M. Owen, J. Petersen, M. Wille, L. Thomsen, A. Dirksen, and M. de Bruijne, "Tree-space statistics and approximations for large-scale analysis of anatomical trees," in *Information Processing in Medical Imaging*, ser. Lecture Notes in Computer Science, 2013, pp. 74–85.

ACKNOWLEDGMENTS

This research was supported by the Lundbeck Foundation; AstraZeneca; The Danish Council for Strategic Research; The Danish Council for Independent Research | Technology and Production; Netherlands Organisation for Scientific Research (NWO); Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation. Megan Owen acknowledges the support of the Fields Institute.