

MACHINE LEARNING FOR LUNG IMAGE ANALYSIS

towards the automatic
quantification of airway diseases

ANTONIO GARCIA-UCEDA JUAREZ



Machine Learning for Lung Image Analysis

towards the automatic
quantification of airway diseases

Antonio Garcia-Uceda Juarez

Acknowledgements:

This research was funded by the Innovative Medicines Initiative (IMI) (n. 115721).

For financial support for the publication of this thesis the following organizations are gratefully acknowledged: Erasmus University Rotterdam and the department of Radiology and Nuclear Medicine of Erasmus MC.

ISBN: 978-94-6458-507-0
Cover: Mathilde Tournebize | www.shutterstock.com
Layout: Antonio Garcia-Uceda Juarez
Printing: Ridderprint | www.ridderprint.nl

© Antonio Garcia-Uceda Juarez, 2022

Except for the following chapters:

Chapter 4: © Springer Nature, 2019

All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without written permission from the author or, when appropriate, from the publisher.

Machine Learning for Lung Image Analysis

towards the automatic
quantification of airway diseases

Machine Learning voor Beeldanalyse van de Longen
richting automatische kwantificering van luchtwegaandoeningen

THESIS

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Wednesday 26 October 2022 at 13.00 hrs

by

Antonio Garcia-Uceda Juarez
born in Madrid, Spain

Doctoral Committee

Promotors Prof. dr. M. de Bruijne
 Prof. dr. H.A.W.M. Tiddens

Other members Prof. dr. S. Klein
 Prof. dr. C.I. Sanchez
 Prof. dr. B. Stoel

This thesis is dedicated to my dad and to my late uncle
Sebastián for being the first examples of courage, perseverance
and dedication that I have known in my life

Esta tesis está dedicada a mi padre y a mi difunto tío Sebastián
por ser los primeros ejemplos de coraje, perseverancia y
dedicación que he conocido en mi vida

Contents

1	Introduction	2
1.1	Respiratory system	2
1.2	Lung and airway diseases	3
1.3	Imaging of the lungs	5
1.4	Segmentation of airways	5
1.5	Automated airway segmentation methods	7
1.6	Automated airway measurement methods	10
1.7	Outline of this thesis	11
2	Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks	16
2.1	Introduction	17
2.2	Methods	19
2.3	Data	22
2.4	Experiments	24
2.5	Results	26
2.6	Discussion	34
2.7	Conclusions	37
2.A	Learning curves of the proposed method	38
2.B	Results grouped by the presence of lung disease	39
2.C	Implementation details of the nnU-Net method	39
3	Label refinement network from synthetic error augmentation for medical image segmentation	44
3.1	Introduction	45
3.2	Methods	47
3.3	Experiments	52
3.4	Results	55
3.5	Discussion	58
3.6	Conclusions	61
4	A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs	66
4.1	Introduction	67
4.2	Methods	67

4.3	Experiments	70
4.4	Results and discussion	72
4.5	Conclusions	74
5	Creating a training set for artificial intelligence from initial segmentations of airways	78
5.1	Introduction	79
5.2	Methods	79
5.3	Results	82
5.4	Discussion	84
5.5	Conclusions	86
5.A	Additional figures and tables	87
6	Reproducibility of a combined artificial intelligence and optimal-surface graph-cut method to automate bronchial parameter extraction	92
7	The effect of CFTR modulators on structural lung disease in cystic fibrosis	96
8	Quantitative chest computed tomography scoring technique for bronchiectasis (BEST-CT)	100
9	Discussion	104
9.1	General discussion	104
9.2	Implications for clinical practice	107
9.3	Recommendations for future research	109
9.4	Conclusion	110
	Summary	113
	Samenvatting	117
	Acknowledgments	121
	About the author	125
	Publications	127
	PhD portfolio	131
	Acronyms	135
	Bibliography	141



CHAPTER 1

Introduction



1.1 Respiratory system

The respiratory system is the group of organs in mammals whose main function is to facilitate oxygen uptake into the bloodstream, and discharge carbon dioxide from the bloodstream into the atmosphere. The respiratory system consists of the upper and lower airways. The nose and pharynx make up the upper airways, while the larynx, trachea, bronchi and bronchioles make up the lower airways. The lungs also contain numerous other structures such as blood vessels to achieve its normal function. The lungs are the main and largest organ of the respiratory system, with a total volume and weight in healthy adults of approximately 6 L and 1 kg [1], respectively. The lungs and their anatomical parts are schematically shown in Figure 1.1.

The human body has a left and a right lung, located and protected inside the thoracic cage. The left lung is smaller than the right lung, as it shares space with the heart. The right lung has three lobes, while the left lung has two lobes. Each lobe is subdivided into bronchopulmonary segments, which are sections supplied with air by separate airways and blood vessels. The lung tissue is a soft spongy material named lung parenchyma, which is composed of the alveolar sacs and connective tissue named interstitium. The alveoli are tiny sacs that fill with air during inspiration, and it is the place where the gas exchange takes place. There are roughly 500 million alveoli in each lung of healthy adults [2]. At the end of inspiration, the lung parenchyma consists of 80–90% air [3]. In the gas exchange process, the oxygen from the inflow of air that fills the alveoli enters the bloodstream, and the carbon dioxide from the bloodstream is released into the alveoli. Then, the resulting air rich in carbon dioxide is discarded during expiration.

The airways or bronchi are tubular structures inside the lungs that conduct the airflow from the larynx into the lungs and inside the alveoli. The airways form a tree-like branching network, named the airway or bronchial tree, where large branches subdivide (or bifurcate) progressively into smaller branches. The first and largest branch is the trachea. This divides into the left and right main bronchi, which supply air into the left and right lung separately. Both main bronchi divide into the lobar (secondary) bronchi, which supply air into separate lung lobes. Then, each lobar bronchus divides into the segmental (tertiary) bronchi, which supply air into separate bronchopulmonary segments. Then, each segmental bronchus divides into the subsegmental bronchi, progressively to branches of 4th, 5th and 6th generation (with respect to the trachea of generation 0). From these bronchi further divisions result in smaller branches called bronchioles. Bronchioles further subdivide many times into branches of smaller and smaller size, until the terminal bronchioles connect to the alveoli. The average number of divisions that airways of healthy adults undergo between the trachea and the alveoli is 23 [4]. Adjacent to the bronchi, there are pulmonary arteries that conduct the bloodstream to the alveoli for the gas exchange. Moreover, there are pulmonary veins that transfer the oxygenated blood from the alveoli to the heart. Both pulmonary arteries and veins also form tree-like branching networks, where large vessels subdivide progressively into smaller vessels. The larger vessels are named similarly to the adjacent bronchi, i.e., as lobar, segmental and

subsegmental arteries and veins. Then, further subdivisions result in smaller vessels called arterioles or venules, which further bifurcate until the terminal capillaries that connect to the alveoli.

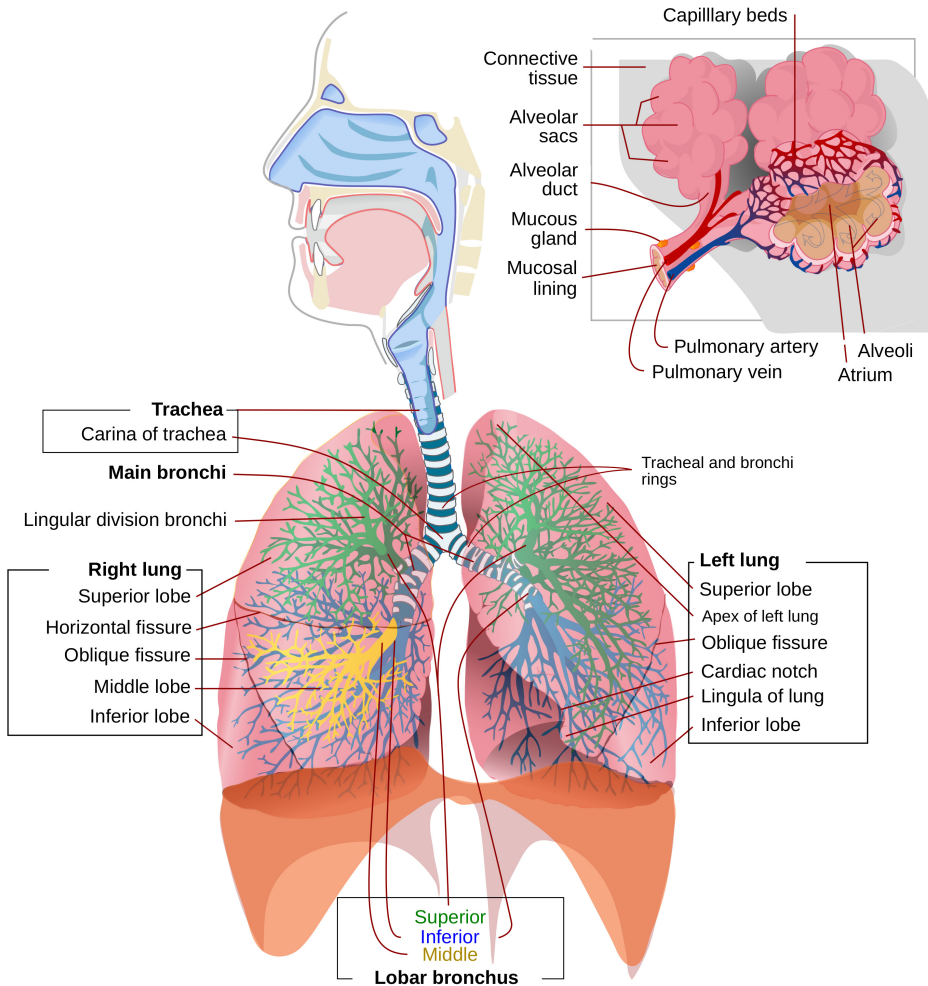


Figure 1.1: Schematic view of the human respiratory system, with the anatomical parts of the lungs. (Adapted from https://en.wikipedia.org/wiki/File:Respiratory_system_complete_en.svg).

1.2 Lung and airway diseases

There are many diseases that affect the lungs and airway tissues and cause structural abnormalities in these organs. Two important lung diseases relevant for the research

presented in this thesis are chronic obstructive pulmonary disease (COPD) and cystic fibrosis (CF). Moreover, a common and important airway condition that characterizes many lung diseases, including COPD and CF, is bronchiectasis, which is also relevant for this thesis.

Bronchiectasis [5, 6, 7] is defined as an abnormal and irreversible enlargement of the airways and thickening of the bronchial wall. The widened airways are the result of chronic inflammation, which weakens the muscle layer and cartilage in the bronchial wall, reducing the airway ability to maintain their shape. The weakened and thickened wall easily folds and airways can close during expiration, causing airflow obstruction. Moreover, the dysfunctional wall reduces the airway ability to clear mucus secretions, which is an important defense mechanism against bacteria, and makes airways more vulnerable to infections. Airway infections can lead to more inflammation and in turn more bronchiectasis, causing a vicious cycle [8]. Two characteristics of bronchiectatic airways are 1) a diameter that is larger than that of the adjacent artery, and 2) a lack of normal tapering, defined as the gradual reduction of the airway diameter along the branch, or between the branches before and after bifurcation.

Chronic Obstructive Pulmonary Disease [7, 9, 10] is an overarching diagnosis that comprises several chronic conditions causing an irreversible obstruction of the airflow and destruction of the lung parenchyma. It is the third leading cause of mortality worldwide [11]. COPD is caused by long-term exposure to toxic gases and particles, such as smoking and to a lesser extent air pollution. COPD is characterized by chronic airway inflammation, bronchiectasis and alveolar destruction, leading to emphysema. Emphysema is defined as an abnormal and permanent enlargement of the air spaces distal to the terminal bronchioles, caused by the weakening and collapse of the alveolar wall. This results in a decrease in the alveolar and the capillary surface area, which reduces the gas exchange. The alveolar destruction also causes a loss of elastic recoil of the lung parenchyma, reducing the lungs ability to exhale. The limited expiratory airflow is exacerbated by the early closure of thickened airways, causing air trapping inside the lungs. Combined with the narrowing of airways due to inflammation, this explains the overall airflow obstruction in COPD.

Cystic Fibrosis [7, 12, 13] is a severe genetic disease characterized by chronic inflammation and infection of the lung and airways, starting from a young age. Although CF affects many organs, lung disease is the major cause of morbidity and mortality. The gene mutation in CF results in a dysfunctional protein (the cystic fibrosis transmembrane conductance regulator or CFTR) that causes the buildup of thick and sticky mucus secretions inside the airways. This excessive mucus facilitates the presence of pathogens, which causes chronic airway infections. These infections lead to airway inflammation and eventually bronchiectasis. The accumulation of mucus is exacerbated by the reduced ability of bronchiectatic airways to clear mucus, which increases the severity of the airway infections. Chronic infections in CF cause progressive structural lung damage, eventually leading to respiratory failure. Another result of the excessive and thick mucus is the formation of mucus plugs inside the airways, which can obstruct especially the small ones.

1.3 Imaging of the lungs

Medical imaging are technologies that allow radiologists to visualize the anatomical structure of human organs for diagnostic purposes. Thoracic computed tomography (CT) is a major imaging technique used to visualize the lungs, as it provides a detailed 3D view of the lung and airway tissues. Chest CT scans consist of a series of cross-sectional images or “slices” of the human thorax, typically ranging between 100 and 600 slices, depending on the size of the scanned subject and the reconstruction protocol. Each CT slice is the result from reconstructing multiple X-ray measurements taken from different angles around the patient, and processing with image reconstruction algorithms. An example of a 3D chest CT scan in different views is shown in Figure 1.2. There is a large variability in the quality of CT scans, depending on the characteristics of the scanner and on the scanning protocols, including 1) the CT resolution (or slice thickness and in-slice voxel size), 2) the dose level and 3) the reconstruction kernel. The dose level refers to the amount of ionizing radiation emitted to take the CT scan, and generally higher dose results in less noisy images. However, CT scanning is performed with the lowest dose needed to obtain diagnostic images, in order to limit the exposure of harmful ionizing radiation to patients. The reconstruction kernel is a parameter of the image reconstruction algorithm, and determines the sharpness of the structures in the CT scan. There are various kernels available, ranging from “smooth” to “sharp”, where generally sharper kernels result in better defined edges with less blurring, but also more noise.

Other thoracic imaging techniques include conventional chest X-rays and magnetic resonance imaging (MRI). Chest X-rays are the most commonly used images, and give a 2D projection of the lungs from the front or side-ways of the human body. The 2D view results in many structures inside the lungs being superimposed, which makes X-rays difficult to interpret and obscures the small airways and vessels. Due to this, chest X-rays are mostly used to detect large or evident lung abnormalities. MRI scans provide a detailed 3D view of the scanned organs and do not use ionizing radiation, which makes them advantageous over CT scans. However, chest MRI scans have had limited use for lung imaging. This is because 1) the lung parenchyma has poor signal intensity on the MRI scan, since the lung is mainly composed of low-density air, and 2) motion artifacts caused by cardiac pulsation and respiration [14]. Due to this, MRI scans for lungs still play a limited role in clinical care in most centers and are mainly limited to vessel imaging and for research purposes.

1.4 Segmentation of airways

The segmentation of the bronchial tree from thoracic CT scans is a useful technique to assess lung diseases that are characterized by structural changes of the airways. By measuring the dimensions of individual airways, such as the diameter and wall thickness, clinicians can accurately quantify the severity and extension of these lung diseases on the CT scan. Moreover, it can provide clinically relevant information about the heterogeneity of the disease. In contrast, other diagnosis techniques such as spirometry only measures the consequence of the lung disease, i.e., the limitation in airflow, and gives limited insight into the underlying causes. Spirometry has also

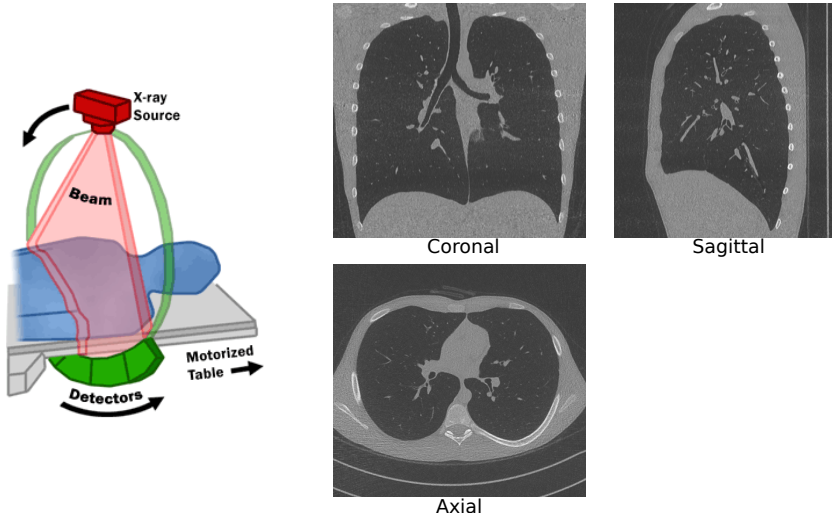


Figure 1.2: Left: Drawing of a CT imaging system, with the CT fan beam applied on the patient. (Sourced from [https://en.wikipedia.org/wiki/File:Drawing_of_CT_fan_beam_\(left\)_and_patient_in_a_CT_imaging_system.gif](https://en.wikipedia.org/wiki/File:Drawing_of_CT_fan_beam_(left)_and_patient_in_a_CT_imaging_system.gif)). Right: Multiplanar view of a low-dose chest CT scan, in coronal, sagittal and axial planes.

limited reproducibility, as the test is difficult to perform correctly, especially for adult or diseased patients.

On the CT scan, the airways appear as hollow tubular structures of many sizes and orientations. The interior of airways (the lumen) has low signal intensity, as it contains air. The airway wall is a thin tissue enclosing the lumen and has high signal intensity. The surrounding background, the lung parenchyma, again has low signal intensity, as it mostly consists of air. The larger bronchi are relatively easy to visualize on the CT scan. However, the bronchioles are challenging to detect and segment manually, with the smaller branches having a cross-sectional area of only a few voxels and very thin walls, often below the scan resolution. An example of the segmented bronchial tree from a chest CT scan is shown in Figure 1.3. The smallest branches that are discernible on the CT scan of an adult subject are up to the 16th generation and have a size of roughly 1–2 mm [15], depending on the body size of the scanned subject. Commonly used low-dose CT scans have low resolution and poor signal-to-noise ratio. This makes the structures inside the lungs appear blurry, and especially the airways whose borders become non-clearly defined. This is further aggravated by image artifacts such as motion effects. Moreover, airways that are severely deformed as a result of severe lung diseases are more difficult to segment. All this makes segmenting the bronchial tree manually from the CT scan an extremely tedious and time-consuming task, taking many hours per scan [16, 17], and susceptible to errors. Thus, an automatic method to segment the bronchial tree is important to quantify airway abnormalities accurately and with little effort. However, the design of a robust and accurate airway segmentation method is challenging, due to 1) the complex 3D morphology of the airway tree and 2) the wide range of CT

characteristics (resolution, dose setting, reconstruction kernel...) that largely influence the performance of the method.

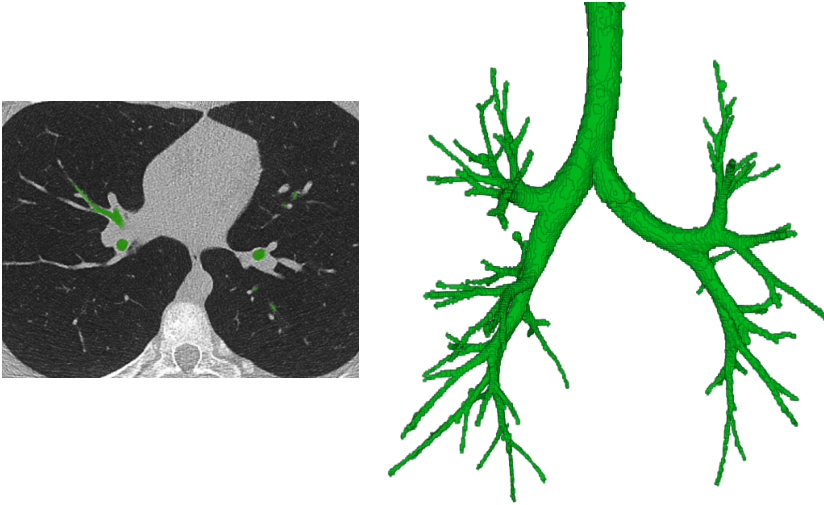


Figure 1.3: Left: Chest CT scan in axial plane overlaid with a segmentation of the airways (in green). Right: 3D view of a segmentation of the full bronchial tree, showing the wide variety of branches present in the tree.

1.5 Automated airway segmentation methods

Several methods have been proposed to perform airway segmentation from chest CT scans. The vast majority of these methods have focused on the segmentation of the airway lumen, while the segmentation of the airway wall has received less attention. An overview of traditional airway segmentation methods is available in [3, 18], and of more recent machine and deep learning based methods in [19]. The first airway segmentation methods [16, 20] used the intensity-based 3D region growing algorithm, relying on the high contrast between the airway lumen, of low intensity, and the airway wall, of high intensity. Region growing is a simple segmentation method that, given an initial seed on the image, selects the neighboring voxels that have an intensity value within a range of user-defined thresholds. The process is iterated until no more neighboring voxels are selected. Region growing can segment accurately the central bronchi, but fails to capture the peripheral airways of smaller size. This mainly occurs as the segmentation leaks into the lung parenchyma near airways with thin and blurred borders, resulting in false positive branches. Moreover, in the presence of mucus plugs in the airways, which have higher signal intensity than the airway lumen, the region growing segmentation will stop prematurely and not detect the remaining branches downstream.

Despite its limitations, intensity-based region growing is a simple algorithm and many methods in the literature are based on it. They typically use region growing

to segment the larger airways, followed by additional approaches to detect smaller airways while preventing leakage [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34]. Other methods use region growing to obtain an initial coarse (leaked) segmentation, followed by an additional filtering approach to remove the false leaked branches [35, 36]. Depending on the nature of these additional operators, the methods can be grouped into 3 categories: 1) morphological based methods, 2) template matching based methods and 3) machine learning based methods. Morphological based methods [21, 22, 23, 24, 25] use morphological operations (dilation, erosion...) designed specifically to detect airways. Template matching based methods [26, 27, 28, 30, 31, 37] use predefined masks or templates, of 2D circular or 3D cylindrical shape and varying sizes, orientations and intensity levels, and test the shape matching of the image data with these masks to identify airways. Machine learning based methods [32, 33, 34, 35, 36] use classifiers that “learn” the underlying probability distribution or patterns in the image that define the airways. These classifiers are optimized using a set of predefined features, such as measures of airway shape, orientation and probability. Additionally, some methods use further rules to detect airways by leveraging their anatomical properties, such as being adjacent to an artery [16, 32]. In the airway extraction challenge EXACT’09 [38], a total of 15 airway segmentation methods were compiled and compared, out of which 11 were region growing based methods.

1.5.1 Deep learning segmentation methods

The state-of-the-art for many biomedical image segmentation tasks are based on deep learning (DL), and in particular convolutional neural networks (CNN) [19, 39, 40]. For airway segmentation, several methods based on CNNs have been proposed since recently [41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55].

CNNs consist of a series of stacked layers that perform operations on pixel array data, taking advantage of the spatial structure in an image. These layers extract relevant image features in a hierarchical manner, with each layer extracting information from the intermediate feature maps output by the preceding layer (or from the input image data for the first layer), and thus deeper layers obtaining more complex and higher-level features. The most characteristic operation of a CNN is convolution, where a kernel of given size is applied on each pixel and its neighbors across the entire input array (feature maps or input image) as a weighted sum of the involved pixel values. Non-linear activation functions (e.g., rectified linear units or ReLU) are typically applied on the convolution output to add non-linearity and enhance the representational power of the CNN. Other characteristic operations of CNNs are 1) pooling, where the number of pixels in the output array is reduced and typically the number of per-pixel features is increased, 2) upsampling, which is the reverse operation of pooling, and 3) dense connections. The output or final layer of the CNN performs the decision making using the highest-level features from the last layer. Depending on the task the CNN is built for, classification, regression or segmentation, the output is a categorical label (e.g., healthy / diseased), a continuous variable, or a pixel map for the target objects, respectively. The layers of a CNN form a set of trainable parameters, i.e., the weights and biases that define the filters in convolutional layers and dense connections in fully connected layers. In supervised learning, these parameters are

optimized by a training algorithm and a cost function that quantifies the differences between the CNN output and the ground truth, so that the parameters are iteratively refined in directions that minimize the cost function.

The layers in CNNs are typically grouped in blocks composed of several stacked convolutional layers (including non-linearity) followed by pooling. With this structure, the CNN extracts features at different lower image resolutions, increasing its receptive field of view from local to more global, and has a reduced number of trainable parameters for the same field of view. In classification or regression CNNs, the latest layers are typically fully connected across all the elements of the input and output arrays. This CNN layout is shown schematically in Figure 1.4. For image segmentation, an improved CNN architecture is the so-called U-Net [56], which is shown schematically in Figure 1.5. This U-Net is composed of two mirror networks, an encoder and a decoder, composed of several blocks of a few convolutional layers followed by pooling or upsampling, respectively; together with skip connections between the encoding and decoding paths at the same resolution. The function of the decoder is to augment the lower-resolution / higher-order features extracted in the encoder with the higher-resolution / lower-order features extracted in earlier layers, linked across the skip connections. Thanks to its design, the U-Net can obtain a pixel segmentation map of the target objects directly from an input image. Currently, the U-Net is the most popular method for airway segmentation, with most U-Net-based approaches in the literature being proposed only in the last few years, including our work in this thesis.

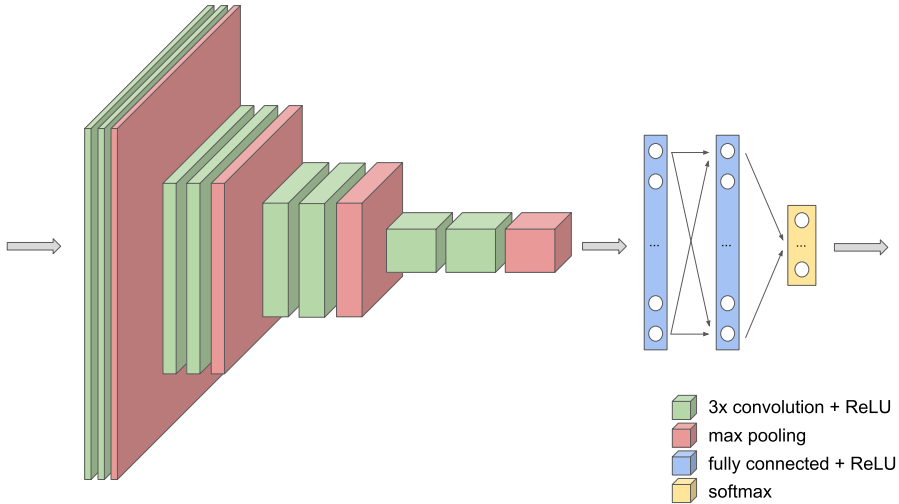


Figure 1.4: Schematics of a convolutional neural network for image classification.

The main advantage of CNNs over classical machine learning methods is that feature extractors in the convolutional layers are learned automatically through data and are end-to-end optimized. In contrast, classical machine learning relies on image features that are predefined or “handcrafted” for a given task, which is complex to do and the resulting features are less powerful than those from CNNs. On the other

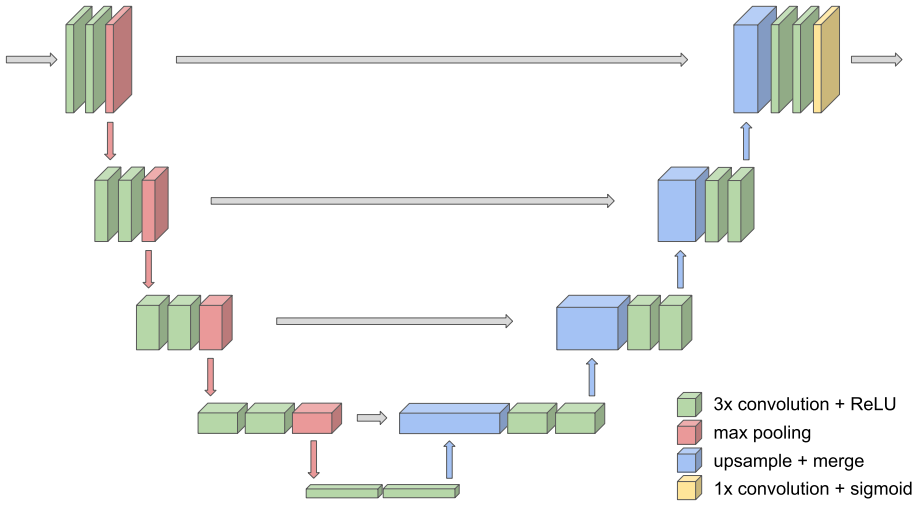


Figure 1.5: Schematics of the U-Net architecture for image segmentation.

hand, CNNs are more difficult to train than classical machine learning and require larger training datasets with ground truth annotations, which are typically tedious to obtain. In fact, it is only since recently that CNNs have become successful thanks to the availability of 1) larger training datasets, 2) improved training algorithms for deep networks, and 3) fast parallel computing in modern GPUs.

1.6 Automated airway measurement methods

Most methods in the literature to obtain airway measures start from a lumen segmentation of the bronchial tree. For measures that require the dimensions of the bronchial wall, the segmentation of the airway outer wall can be extracted from that of the lumen [57, 58, 59, 60]. As a first step, the airway centerline tree is extracted from the lumen segmentation [38], and single branches are detected as centerline segments between two bifurcation points, or between a bifurcation and the end of terminal branches. Then, several airway dimensions are extracted for every branch, such as the diameters of the lumen and outer wall, the cross-sectional area and the wall thickness. Due to the variation of the airway cross-sectional shape across a branch, these airway dimensions are typically computed by averaging a set of local dimensions extracted at regular intervals along the centerline [61, 62]. The branch generation is also extracted, which can be defined as the number of branch bifurcations counted in the path linking the given branch and the root of the airway tree, i.e., the trachea [63].

The extracted airway dimensions are used to compute airway biomarkers, such as airway tapering (intra-branch and inter-branch), airway-artery ratio, wall-artery ratio, wall area percentage and Pi10. These biomarkers are defined as follows:

1. *Intra-branch tapering*: the percentage reduction of the airway diameter (lumen

or outer wall) per millimeter along the branch centerline.

2. *Inter-branch tapering*: the percentage ratio of the difference between the airway diameter (lumen or outer wall) and that of the branch before bifurcation (parent), to the diameter of the parent airway.
3. *Airway-artery ratio*: the ratio between the airway diameter (lumen or outer wall) and the diameter of the adjacent artery.
4. *Wall-artery ratio*: the ratio of the difference between the airway outer wall and lumen diameters, to the diameter of the adjacent artery.
5. *Wall area percentage*: the percentage ratio between the cross-sectional areas of the airway wall and the whole airway.
6. *Pi10*: the square root of the airway wall area estimated for a fictitious airway with 10 mm of lumen diameter, obtained by projecting the set of extracted dimensions using linear regression [64].

The airway intra-branch [65, 66, 67] and inter-branch [65] tapering has been used to quantify bronchiectasis. The airway-artery and wall-artery ratios [17, 62, 68, 69, 70, 71] have been extensively used to quantify the bronchiectasis and airway wall thickening due to CF lung disease. The airway wall thickness [61, 72], wall area percentage [72] and Pi10 [59, 72, 73] have been used to quantify the airway narrowing and wall thickening due to COPD. The airway tapering, wall thickness and wall area percentage depend on the size of the airways, and therefore cannot be compared across all airways in the lungs. Instead, these biomarkers are summarized and compared per airway generation, since airways of the same generation have similar sizes and characteristics. The Pi10 overcomes this limitation by using a projected wall area to a fixed airway size. However, this biomarker does not quantify real individual branches. The airway-artery and wall-artery ratios are less dependent on the airway size since they use as normalization factor the diameter of the adjacent artery. This is because an airway and its accompanying artery have similar sizes in healthy subjects [74]. However, these biomarkers require additionally the dimensions of the vascular tree in the lungs, and to identify airway-artery pairs. To do this, a segmentation of the vessel tree can be obtained [32] and airways and arteries can be paired based on similarity in size, proximity and orientation [62].

1.7 Outline of this thesis

This thesis focuses on developing automatic image processing methods to segment the bronchial tree from chest CT scans and to subsequently extract airway measures as imaging biomarkers. The contributions and outline of this thesis are as follows:

1. Chapter 2 presents a fully automatic method to segment the bronchial tree, based on the deep learning U-Net architecture. The method is validated on three datasets, including challenging CT scans with airway abnormalities. The results show that this method obtains high-quality and complete airway segmentations,

and outperforms several other deep learning airway segmentation methods. This method is implemented in an open-source software and made available in a public repository.

2. Chapter 3 presents a novel method to correct errors in the initial segmentations from a U-Net. This method uses synthetic labels augmented with realistic errors, and is trained to correct these errors in the segmentation results. The method is validated for the segmentation of lung airways and brain vessels. The results show that this method can correct errors and obtain more complete segmentations.
3. Chapter 4 presents a novel segmentation method that combines the U-Net with a graph neural network model. The graph model allows the network to extract information from a larger region of the images, which can help improve the segmentation decisions. The method is validated for airway segmentation. The results show that this method can obtain slightly more complete airway segmentations, with slightly more errors.
4. Chapter 5 presents an approach to efficiently generate ground truth airway segmentations. This consists of manually correcting initial airway segmentations obtained from a pre-trained U-Net. The approach is validated by correcting the initial airway segmentations from a small dataset, and using these as ground truth to retrain the U-Net. The results from the retrained model were more accurate and complete than the initial segmentations.
5. Chapter 6 presents an automated pipeline to segment both the lumen and wall surfaces of the bronchial tree. It combines a U-Net for airway extraction with an optimal-surface graph-cut method that segments the lumen and wall around the extracted airways. This approach is validated by measuring bronchial parameters from the segmentations and assessing their repeatability. The results show that the measurements were adequately reproducible.
6. Chapter 7 presents a study of the effect of CFTR modulator therapy on structural lung abnormalities in patients with cystic fibrosis lung disease. We used three different CT analysis methods: a quantitative method based on the median of airway-artery ratio and airway tapering measures extracted over the small airways, and two visual scoring systems, one quantitative and one semi-quantitative. The results show that quantitative methods were more sensitive to detect improvements in structural lung disease due to therapy.
7. Chapter 8 presents a quantitative CT analysis method to assess structural lung abnormalities in patients with non-cystic fibrosis bronchiectasis, named BEST-CT. We compared BEST-CT with two other methods: a quantitative method based on the median of luminal intra-branch tapering measures extracted over all the airways, and a semi-quantitative visual scoring system. The results show that the BEST-CT method can capture the heterogeneity of bronchiectasis, and correlates well with the other methods.
8. Chapter 9 provides a general discussion of the findings and achievements of the research presented in this thesis, and gives possible directions for future research.



CHAPTER 2

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



Abstract

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

Based on: **A. Garcia-Uceda**, R. Selvan, Z. Saghir, H. Tiddens, and M. de Bruijne, "Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, p. 16 001, 2021. DOI: [10.1038/s41598-021-95364-1](https://doi.org/10.1038/s41598-021-95364-1)

2.1 Introduction

Segmentation of the airway tree from thoracic computed tomography (CT) scans is a useful procedure to assess pulmonary diseases characterized by structural abnormalities of the airways, such as bronchiectasis (widening of the airways) and thickening of the airway wall. To quantitatively assess these conditions on the CT scan, clinicians are interested in having individual airway measurements, including airway diameter, wall thickness and tapering [17, 65]. The bronchial tree is a complex 3D structure, with many branches of varying sizes and different orientations. Segmenting airways by manual or semi-automatic methods is extremely tedious and time-consuming, taking more than 15 hours [17] per scan, manually, or up to 2.5 hours [75] per scan, semi-automatically. Thus, a fully automatic airway segmentation method is important to provide an accurate, effortless and free of user-bias segmentation of the airway tree needed to quantify airway abnormalities.

In CT scans, airways appear as tubular structures with an interior of typically low intensity (the lumen) enclosed by a structure of higher intensity (the airway wall). Moreover, airways are surrounded by a background which can be of low intensity (the lung parenchyma) or higher intensity (typically the vessels). The earliest airway segmentation methods used the region growing algorithm [16, 20] to segment the airway lumen, relying on the intensity difference between the airway lumen and wall. Region growing can accurately capture the central bronchi, but has a tendency to systematically fail at extracting the peripheral bronchi of smaller size, missing a large portion of the airway tree. Also, when segmenting the smaller airways it often results in segmentations leaking into the lung parenchyma. This is due to the reduced intensity difference and reduced signal-to-noise ratio between the airway lumen and wall, caused by partial volume effects near the smaller airways. Many airway segmentation methods build upon the region growing algorithm, using this as an initial step to segment the larger bronchi and then apply additional operators to detect smaller airways while preventing leakage [24, 25, 30, 32, 33]. Extensions to region growing based methods have been widely reviewed [3, 18]. In the airway extraction challenge EXACT'09 [38], 15 algorithms were evaluated and compared, out of which 11 were region growing based methods. The results showed that all participating methods missed a significant amount of the smaller branches, and many methods had a large number of false positives errors.

Several airway segmentation methods use machine learning classifiers, either for voxelwise airway classification [32, 34] or to remove false positive airway candidates from a leaky baseline segmentation [35, 36]. These classifiers (k -nearest neighbors (kNN) [32], AdaBoost [35], support vector machines [36] or random forest [34]) operate on a set of predefined image features: multiscale Gaussian derivatives [32], multiscale Hessian-based features [34, 35, 36] or image texture features with local binary patterns [34]. These methods can achieve more complete airway tree predictions than previous purely intensity-based methods, with fewer false positives. However, they are highly dependent on the image features used to train the classifier, and may be time-consuming to apply due to the time needed to compute the image features [32, 34, 36]. Recently, many state-of-the-art methods for medical image segmentation tasks have used deep learning [40], and in particular convolutional neural networks (CNN) [76]. The

main advantage of deep CNN methods over classical learning-based techniques is that the extraction of relevant image features is done automatically from data in an end-to-end optimized setting. Several CNN-based methods have been applied for airway segmentation [42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 77]. Charbonnier et al. [43] applied CNNs to detect and remove leakage voxels from a leaked region growing based segmentation. Yun et al. [44] applied the so-called 2.5D CNN, a pseudo 3D CNN which processes the three perpendicular 2D slices around each voxel, to perform voxelwise airway classification. The methods [42, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 77] are based on the U-Net architecture [56]. This model and its 3D extension [78, 79] are highly successful for medical image segmentation. The main advantage of the U-Net over voxelwise CNN models is that it can process entire images in one forward pass through its encoding / decoding structure, resulting in a segmentation map directly. Jin et al. [45] and Meng et al. [46] applied the U-Net on local volumes of interest around airways, being guided by the centerlines from a baseline segmentation [45] or by tracking the airways extracted from the U-Net [46]. Garcia-Uceda et al. [77] applied the U-Net on large image patches extracted from the 3D CT scans, using various data augmentation techniques. Garcia-Uceda et al. [42] proposed a joint approach combining both 3D U-Net and graph neural networks (GNN) [80]. Qin et al. [47, 48] applied the U-Net for 26-fold prediction of the 26-neighbor connectivities of airway voxels, and segments airways by grouping voxels with at least one mutually predicted connectivity with a neighbor voxel. Zhao et al. [49] combined both 2D and 3D U-Nets with linear programming-based tracking to link disconnected components of the segmented airway tree. Wang et al. [50] used the U-Net with spatial recurrent convolutional layers and a radial distance loss to better segment tubular structures. Qin et al. [51, 52] extended the U-Net with feature recalibration and attention distillation modules that leverage the knowledge from intermediate feature maps of the network. Zheng et al. [53] proposed a “General Union” loss to alleviate the intra-class imbalance between large and small airways. Zhou et al. [54] extended the U-Net with a multi-scale feature aggregation module based on dilated convolutions, to include more contextual information from a larger receptive field. Nadeem et al. [55] used the U-Net followed by a freeze-and-grow propagation algorithm to iteratively increase the completeness of the segmented airway tree.

In this work, we present a fully automatic and end-to-end optimized method to perform airway segmentation from chest CT scans, based on the U-Net architecture. A preliminary version of this work was presented in [77]. The proposed method processes large 3D image patches, often covering an entire lung, in a single pass through the network. We achieve this by using a simple U-Net backbone with low memory footprint, and having efficient operations that feed image data to the network. This makes our method simple, robust and efficient. We performed a thorough validation of the proposed method on three datasets with very different characteristics, from subjects including pediatric patients and adults and with diverse airway abnormalities, including the EXACT’09 public dataset [38]. We compared the method with other state-of-the-art segmentation methods tested on the same data. On the EXACT’09 data, we compared with all the recent methods in the literature evaluated on these data, in terms of the average performance measures reported by the EXACT’09 evaluation. On the other datasets, we compared with the two methods previously evaluated on the

same data [32, 33]. Moreover, we evaluated the accuracy of our method with respect to the presence of lung disease and airway abnormalities in the CT scans from the three datasets.

This chapter is organized as follows: in Section 2.2 we explain the proposed methodology for airway segmentation, in Section 2.3 we present the data used for this work, in Section 2.4 we explain the experiments performed to validate the method, in Section 2.5 we present the results obtained, in Section 2.6 we discuss these results and explain the advantages and limitations of the proposed method, and in Section 2.7 we give the conclusions.

2.2 Methods

2.2.1 Network architecture

The network consists of an encoder (downsampling) path followed by a decoder (upsampling) path, at different levels of resolution. Each level in the downsampling or upsampling path is composed of two $3 \times 3 \times 3$ convolutional layers, each followed by a rectified linear (ReLU) activation, and a $2 \times 2 \times 2$ pooling or upsampling layer, respectively. After each pooling or upsampling layer, the number of feature channels is doubled or halved, respectively. There are skip connections between opposing convolutional layers prior to pooling and after upsampling, at the same resolution level. The final layer is a $1 \times 1 \times 1$ convolutional layer followed by a sigmoid activation function. This network is schematically shown in Figure 2.1.

For analysis of 3D chest CT scans, we found that the main bottleneck is the memory footprint of the network. To alleviate this, we use non-padded convolutions in the first 3 resolution levels of the U-Net, where the outmost layers of voxels in the feature maps are progressively removed after each convolution. We still use zero-padding in the remaining levels after the third, to prevent an excessive reduction of the output size of the network. This allows to reduce the memory footprint by approximately 30% compared to a model with zero-padding in all layers. Moreover, we do not use dropout or batch normalization layers, as these require extra memory to store the feature maps after the operation.

2.2.2 Training of networks

The network is trained in an end-to-end supervised fashion, using the CT scans as input and evaluating the network output with respect to the ground truth, the reference airway segmentations, using the soft Dice loss [81],

$$\mathcal{L} = 1 - \frac{2 \sum_{x \in N_L} p(x)g(x)}{\sum_{x \in N_L} p(x) + \sum_{x \in N_L} g(x)} \quad 2.1$$

where $p(x), g(x)$ are the predicted voxelwise airway probability maps and airway ground truth, respectively. In the loss computation, the ground truth is masked to the region of interest (ROI), the lungs, indicated by the sub-index L , so that only voxels within this region contribute to the loss. This mask removes the trachea and part of

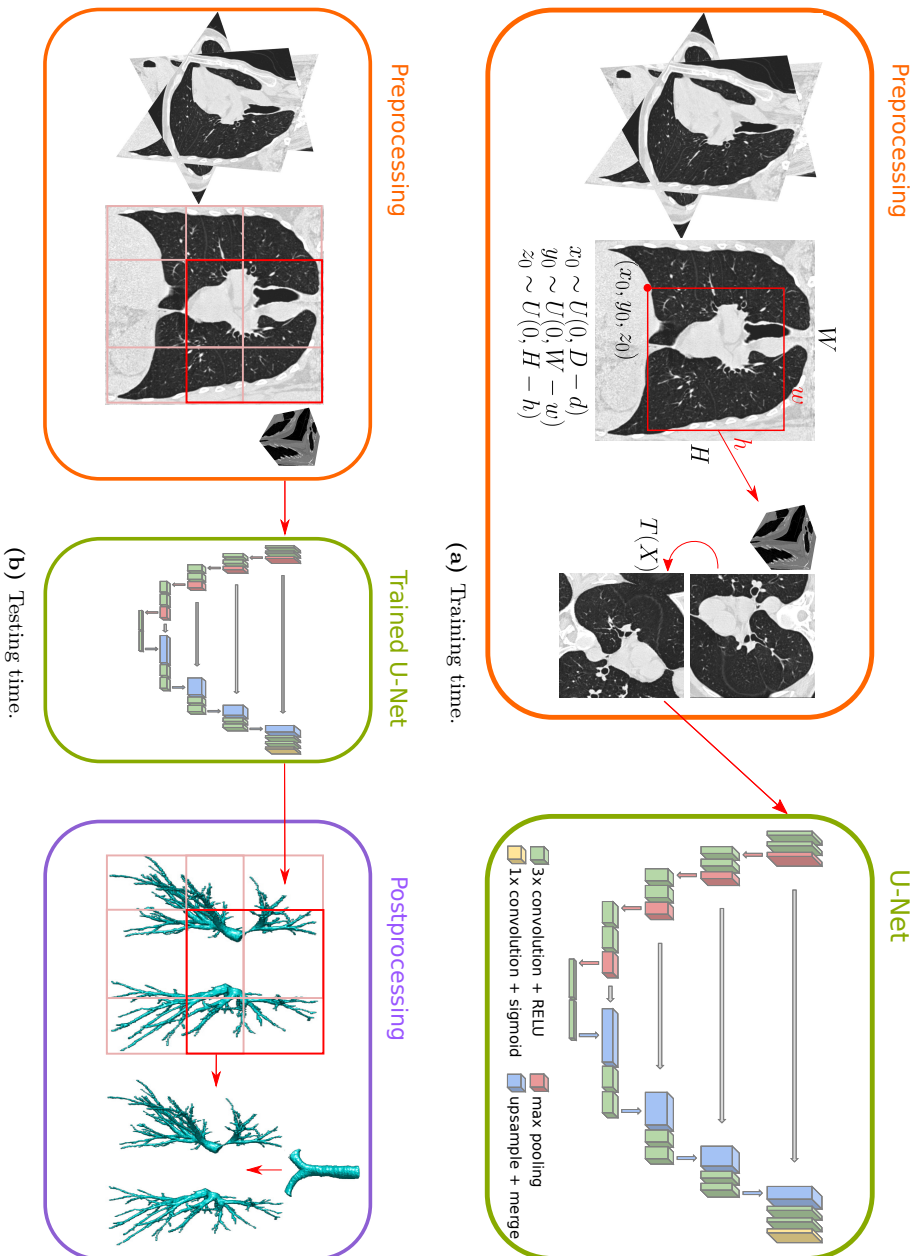


Figure 2.1: Schematics of the proposed airway segmentation method. (a) Overview at training time, showing the extraction of 3D random patches from the CT scan, to feed as input to the U-Net. (b) Overview at testing time, showing the extraction of 3D patches through sliding-window, and the generation of the airway tree segmentation from the patch-wise output of the U-Net.

the main bronchi from the ground truth, so that the training of the network focuses on the smaller branches. The lung segmentation needed for this masking operation is easily computed by a region growing method [32].

2.2.3 Implementation of networks

When building the U-Net, the degrees of freedom (or hyperparameters) that determine the capacity of the network are 1) the number of resolution levels in the U-Net, 2) the number of feature maps in the first layer, 3) the input image size, and 4) the training batch size. We optimized these hyperparameters using subsets of training data from the two datasets used in our experiments. We observed that the input image size was the most important parameter, and that by using larger input patches we obtained better results and faster convergence of the training and validation losses. With this, we found an optimal U-Net of 5 levels, 16 feature maps in the first layer, input size of $252 \times 252 \times 252$, and training batches containing only one image. This network can fit in a GPU NVIDIA Quadro P6000 with 24 GB memory during training. This U-Net construction is used for all the experiments undertaken. The implementation of the network is done using the Pytorch framework [82].

2.2.4 Generation of images during training

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

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

3. Apply random rigid transformations to the input image patches as data augmentation. These transformations include 1) random flipping in the three directions, 2) random small 3D rotations up to 10 degrees, and 3) random scaling with a scale factor in the range 0.75–1.25. The same transformations are applied to the ground truth using nearest-neighbor interpolation.

This series of operations is schematically shown in Figure 2.1. We implemented the patch extraction (2.) and data augmentation (3.) operations in an efficient on-the-fly image generation pipeline that feed data to the network, to reduce any computational overhead for the batch data loading part of the training algorithm. The initial cropping (1.) operation is applied only once to the chest CT scans and ground truth and prior to the training.

2.2.5 Airway extraction

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

1. Aggregate the output patches from the network with airway probability maps by stitching them together (i.e., reversing the sliding-window operation used to extract the input patches). To account for the overlap between patches, we divide the aggregated voxelwise airway probabilities by the number of patches containing the given voxel.
2. Mask the full-size airway probability map to the mask of the lung fields. This is to discard the noise predictions outside the lung regions, as only these regions are included in the training of the networks by Equation 2.1.
3. Threshold the airway probability maps to obtain the airway binary segmentation, using a threshold of 0.5 by default. This output does not contain the trachea and part of the main bronchi that were outside the mask of the lung fields. To include this, we merge the airway segmentation with a mask for the trachea and main bronchi. This is easily computed by a region growing method [32], and then masking its output to the mask of the lung fields.
4. Apply connected component analysis [83] to the airway segmentation and retrieve the final airway tree as the largest 26-neighbor connected component.

2.3 Data

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

1. **CF-CT**: This dataset consists of 24 inspiratory low-dose chest CT scans from pediatric patients at Erasmus MC-Sophia, 12 controls and 12 diseased [17]. The 12 controls were patients with normal lung assessment from CT reported by two different radiologists, who received CT scanning for diagnostic purposes. The 12 diseased were 11 patients with cystic fibrosis (CF) lung disease and 1 with common variable immune deficiency (CVID), who had structural lung damage with airway abnormalities visible on the CT scan. The two groups were gender and age matched, with the age range from 6 to 17 years old in both groups and 5 females per group. Scanning was undertaken using spirometry-guidance in a Siemens SOMATOM Definition Flash scanner. Each CT scan has an in-plane voxel size in the range 0.44–0.71 mm, with slice thickness between 0.75–1.0 mm, and slice spacing between 0.3–1.0 mm.
2. **DLCST**: This dataset consists of 32 inspiratory low-dose chest CT scans from the Danish Lung Cancer Screening Trial [84]. Participants to the trial were subjects between 50 to 70 years old, with a history of at least 20 pack-years of smoking, and without lung cancer related symptoms. Roughly half of the CT scans are from subjects with chronic obstructive pulmonary disease (COPD), reported from spirometry measures. Scanning was undertaken using a MDCT scanner (16 rows Philips Mx 8000, Philips Medical Systems, Eindhoven, The Netherlands). All CT scans have a voxel resolution of roughly $0.78 \times 0.78 \times 1 \text{ mm}^3$.
3. **EXACT’09**: This is the multi-center, public dataset of the EXACT’09 airway extraction challenge [38]. We used for evaluation purposes the EXACT’09 test set consisting of 20 chest CT scans. The conditions of the scanned patients vary largely, ranging from healthy volunteers to subjects with severe lung disease. The data includes both inspiratory and expiratory CT scans, ranging from clinical to low-dose. The CT scans were acquired with several different CT scanner brands and models, using a variety of scanning protocols and reconstruction parameters. Each CT scan has an in-plane voxel size in the range 0.55–0.78 mm, with slice thickness between 0.5–1.0 mm.

2.3.1 Generation of reference segmentations

Networks were trained and evaluated using airway lumen segmentations that were obtained by a combination of manual interaction and automated surface detection. For the CF-CT data, we have centerline annotations manually drawn by trained experts. For the DLCST data, we have airway extractions obtained from the union of the results of methods [32] and [33], and corrected by a trained observer. The visual assessment was done similarly to the EXACT’09 challenge [38], where the obtained airways trees were divided in branches using a wave front propagation algorithm that detects bifurcations, and spurious branches were removed manually. “Partly wrong” branches whose segmentation covered the airway lumen but leaked outside the airway wall were also removed. To obtain the ground truth airway segmentations, we applied the optimal-surface graph-cut segmentation method [58] on top of these initial airway references in order to 1) determine accurate lumen contours and 2) homogenize the ground truth for the two datasets. To evaluate the networks, we used these two

ground truth segmentations as well as (via submission to the EXACT'09 challenge) the reference segmentations from the EXACT'09 test set. To build the EXACT'09 reference, the airway trees predicted by all 15 participating methods to the challenge were evaluated by independent expert observers, and the group of correct-scored branches were merged together.

2.4 Experiments

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

1. **CF-CT**: We compared the performance of our method with that of a previously published airway segmentation method on these data [62]. These results were obtained by a kNN-classifier together with a vessel similarity measure [32], and refined with an optimal-surface graph-cut method [58] to obtain an accurate lumen segmentation. We denote this method by **kNN-VS**. We trained and evaluated our U-Net model using the same 6-fold cross-validation setting as in [62]. The training and testing data in each fold have an equibanced number of CT scans from control and diseased subjects.
2. **DLCST**: We compared the performance of our method with that of two previously published airway segmentation methods on these data [32] and [33]. The method [32] is a kNN-classifier together with a vessel similarity measure, while the method [33] extends the airways iteratively from an incomplete segmentation with locally optimal paths. We refined these results [32, 33] with an optimal-surface graph-cut method [58] to obtain a more accurate lumen segmentation, and to homogenize them with the training data used to train our model, for a fair comparison. We denote these methods by **kNN-VS** and **LOP**, respectively. We trained and evaluated our U-Net model using the same 2-fold cross-validation setting as in [32] and [33]. The training and testing data in each fold have a similar number of CT scans from healthy and diseased subjects.
3. **EXACT'09**: We compared the performance of our method with that of the 15 methods participating in the challenge EXACT'09 based on the results reported in [38], 6 post-challenge methods evaluated on the EXACT'09 data and reported in [85], and 4 recent airway segmentation methods evaluated on these data [43, 44, 52, 86]. The methods in Charbonnier et al. [43], Yun et al. [44] and Qin et al. [52] were previously described in Section 2.1. The method in Gil et al. [86], named PICASSO, uses locally adaptive optimal thresholds learned from a graph-encoded measure of the airway tree branching. We denote these methods by **CNN-Leak**, **2.5D CNN**, **U-Net-FRAD** and **PICASSO**, respectively. The method in Zheng et al. [53] is also evaluated on EXACT'09 data, but we did not include it in our comparison as the authors did not report the same specificity metric (false positive rate) as in the EXACT'09 evaluation. To our knowledge, these are all the recent methods in the literature that are evaluated on the EXACT'09 data. We trained our U-Net model using half of the

CT scans from the CF-CT and DLCST datasets (28 scans in total). This training data has a similar number of CT scans from healthy and diseased subjects. We did not include the EXACT'09 training set in our training data as there are no reference segmentations available. Additionally, we compared our method with the nnU-Net segmentation method proposed by Isensee et al. [87], applied to airway segmentation. We denote this method by **nnU-Net**. To do this, we performed an experiment with the nnU-Net model. We trained it using the same training data as for our method, and we evaluated it on EXACT'09 data following the same evaluation protocol. We provide implementation details of the nnU-Net method and of our experiment in the Appendix.

For each cross-validation fold in the experiments, we used roughly 80% of images in the training fold to train the models, and the remaining 20% is used to monitor the validation loss. To train the models, we used the Adam optimizer [88] with an initial learning rate of 1×10^{-4} , which was chosen as large as possible while ensuring convergence of the training and validation losses. As convergence criterion, we monitored the moving average of the validation loss over 50 epochs, and training is stopped when its value 1) increases by more than 5%, or 2) does not decrease more than 0.1% over 20 epochs. We trained the models until convergence is reached, and we retrieved the model with the overall minimum validation loss for testing. Training time was approximately 1–2 days on a GPU NVIDIA Quadro P6000, while test time inference was less than 1 min per scan including all post-processing steps to obtain the binary airway tree.

To compute the airway predictions on the EXACT'09 data, we thresholded the airway probability maps with a value of 0.1, and retrieved the final airway tree as the largest 6-neighbor connected component. The lower threshold was to compensate for the reduction of completeness in our results when computing a single 6-connected structure, required by the submission guidelines to the EXACT'09 challenge [38]. In contrast, for the predictions on the CF-CT and DLCST data, we used the default threshold of 0.5 and 26-neighbor connected component analysis.

2.4.1 Evaluation

To compare the different methods on the CF-CT and DLCST data, we computed 1) tree length detected, 2) centerline leakage and 3) Dice coefficient, with respect to the ground truth segmentations. For the results on the EXACT'09 data, we evaluated 1) tree length detected and 2) false positive rate, computed by the EXACT'09 challenge organizers [38]. Moreover, we assessed the segmentation accuracy of our U-Net-based method with respect to the presence of lung disease in the CT scans from each dataset. For this, we computed 1) tree length detected, 2) centerline leakage (or false positive rate on the EXACT'09 data) and 3) total tree length detected, and compared the measures from both healthy and diseased subjects.

The metrics are defined as:

1. *Tree length detected (TL)*: The number of ground truth centerline voxels that are inside the predictions, as a percentage of the ground truth centerline full length.

2. *Centerline leakage (CL)*: The number of predicted centerline voxels outside the ground truth, as a percentage of the ground truth centerline full length.
3. *False positive rate (FPR)*: The number of false positive voxels outside the ground truth, as a percentage of the total number of ground truth voxels.
4. *Dice similarity coefficient (DSC)*: The voxelwise overlap between the predictions and the ground truth according to Equation 2.2, with P and G the binary prediction and ground truth masks, respectively.

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad 2.2$$

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

The trachea and main bronchi are removed in these measures from both the predictions and ground truth, similar to [38]. The centerlines are obtained by applying skeletonization [89] to the prediction and ground truth masks.

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

2.5 Results

2.5.1 Evaluation on CF-CT data

We show in Figure 2.2 the evaluation on the CF-CT data of our U-Net-based method and the kNN-VS method [62]. Comparing the U-Net with the kNN-VS results, the former shows higher TL (83.5 (80.7–87.1) vs. 70.1 (58.9–73.9), $p<0.001$), higher DSC (0.876 (0.854–0.883) vs. 0.806 (0.711–0.839), $p<0.001$) and a similar CL (6.09 (4.41–13.6) vs. 9.58 (5.98–22.1), $p=0.021$). This indicates that our U-Net-based method predicts more complete airway trees than the kNN-VS method [62], with more and/or longer peripheral branches, and with a similar count of false positive errors. We show in Figure 2.3 visualizations of airway trees obtained by the two methods. We can see that the airways predicted by our U-Net-based method match better the reference segmentations, with more and/or longer peripheral branches detected, while the kNN-VS method misses many of these peripheral branches.

Comparing the measures from our U-Net-based method between control and subjects with CF and airway abnormalities, we found no significant differences in TL (84.1 (78.6–87.9) vs. 83.5 (81.7–86.4), $p=0.87$) and in CL (5.74 (3.91–13.8) vs. 6.45 (4.68–11.2), $p=0.78$), but they were significant in total tree length (146 (135–156) vs. 263 (208–308), $p<0.01$). This indicates that our method has similar accuracy with

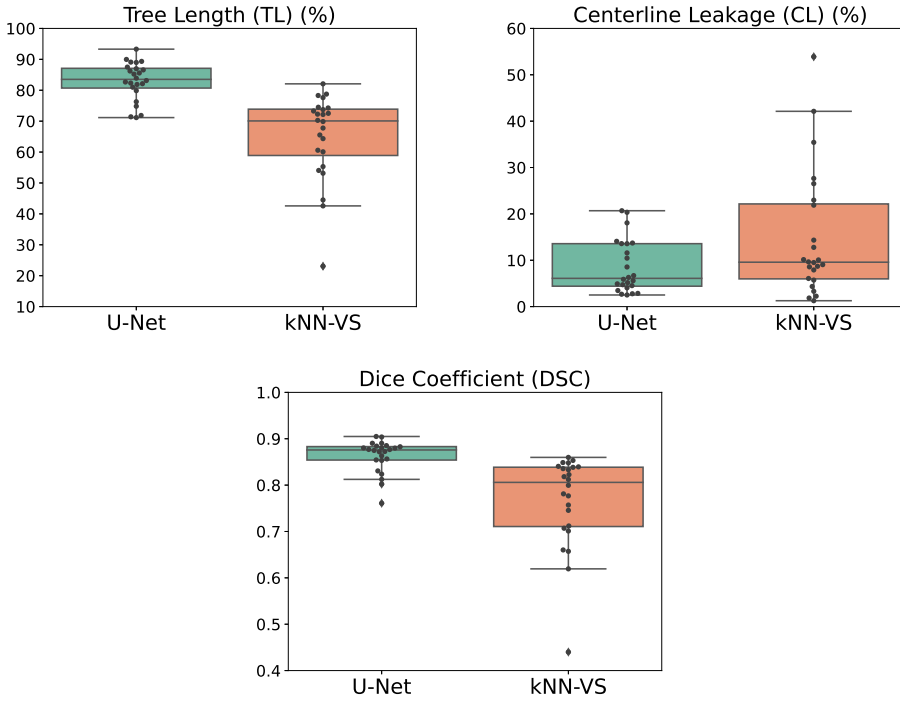


Figure 2.2: Boxplots showing 1) tree length detected, 2) centerline leakage and 3) Dice similarity coefficient on the CF-CT data, for the results obtained by our U-Net-based method and the kNN-VS method [62]. For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show all the data points.

respect to the manual annotations on CT scans from both control and subjects with CF, while segmenting more airway branches in the diseased cases. We show these grouped measures in Figure 2.B.1a in the Appendix.

2.5.2 Evaluation on DLCST data

We show in Figure 2.4 the evaluation on the DLCST data of our U-Net-based method and the kNN-VS [32] and LOP [33] methods. Comparing the U-Net with the kNN-VS results, the former shows lower CL (8.25 (6.26–9.58) vs. 12.0 (9.68–14.9), $p < 0.001$), higher DSC (0.916 (0.909–0.924) vs. 0.810 (0.787–0.825), $p < 0.001$) and a similar TL (81.5 (79.2–84.1) vs. 80.3 (76.1–84.4), $p = 0.19$). Comparing the U-Net with the LOP results, the former shows lower CL (8.25 (6.26–9.58) vs. 21.0 (16.0–25.9), $p < 0.001$), higher DSC (0.916 (0.909–0.924) vs. 0.795 (0.780–0.812), $p < 0.001$), but lower TL (81.5 (79.2–84.1) vs. 96.8 (94.9–97.7), $p < 0.001$). This indicates that our U-Net-based method predicts an airway tree as complete as the kNN-VS method [32], and less

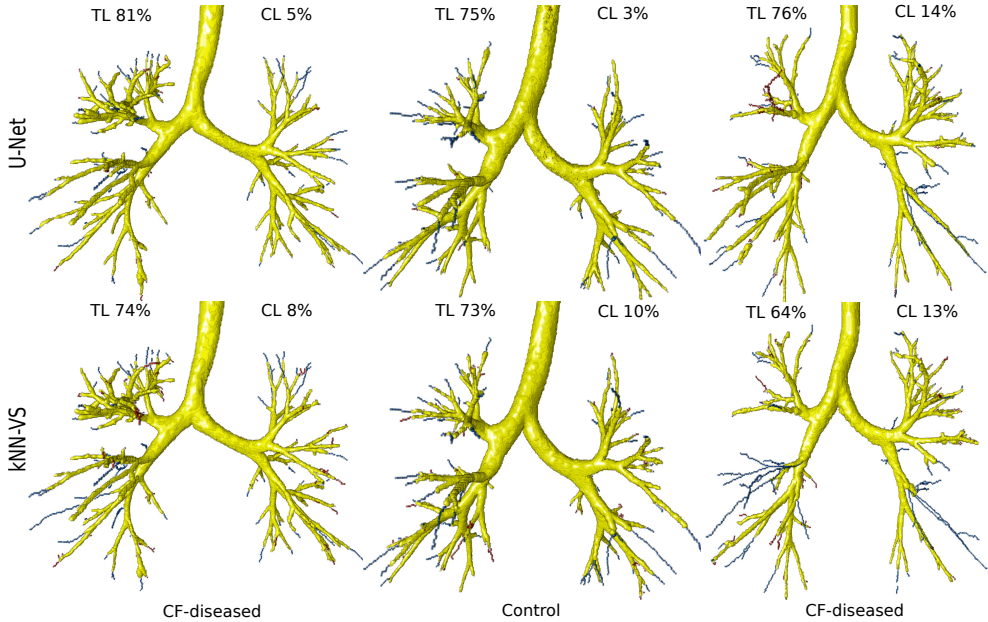


Figure 2.3: Visualization of predicted airways trees on the CF-CT data, obtained by our U-Net-based method and the kNN-VS method [62]. True positives are displayed in yellow, centerline false negatives in blue and centerline false positives in red. These false negatives and positives are the errors counted in the tree length detected and centerline leakage measures reported in Figure 2.2, respectively. The cases displayed correspond to, from left to right, the two U-Net results with Dice score closest to the median of the test Dice measures, and the U-Net result with the lowest Dice score.

complete than the LOP method [33], but with significantly less false positive errors than these two methods. It should be noted that this comparison is biased towards the kNN-VS and LOP methods that made up the reference segmentations, which partly explains the very high completeness of the results by the LOP method. We show in Figure 2.5 visualizations of airway trees obtained by the three methods. We can see that the airways predicted by the LOP method have more peripheral branches detected (with almost no false negative errors), however they have more false positive errors. The results from our U-Net-based method have less peripheral branches detected, but also have much less false positives.

Through visual inspection of the results, we observed that several false positive errors correspond to real airways on the CT scan that were missing in the reference segmentations. This is because this reference was built in a conservative way from automatic airway extractions, where branches not extracted by either method [32] or [33] were not included, and branches evaluated as “partly wrong” were removed, as explained in Section 2.3.1. Thus, the reference segmentations are somewhat incomplete. For the segmentation results obtained by the three methods, we inspected any false positives that were of tubular shape and that were long enough to be clearly perceived

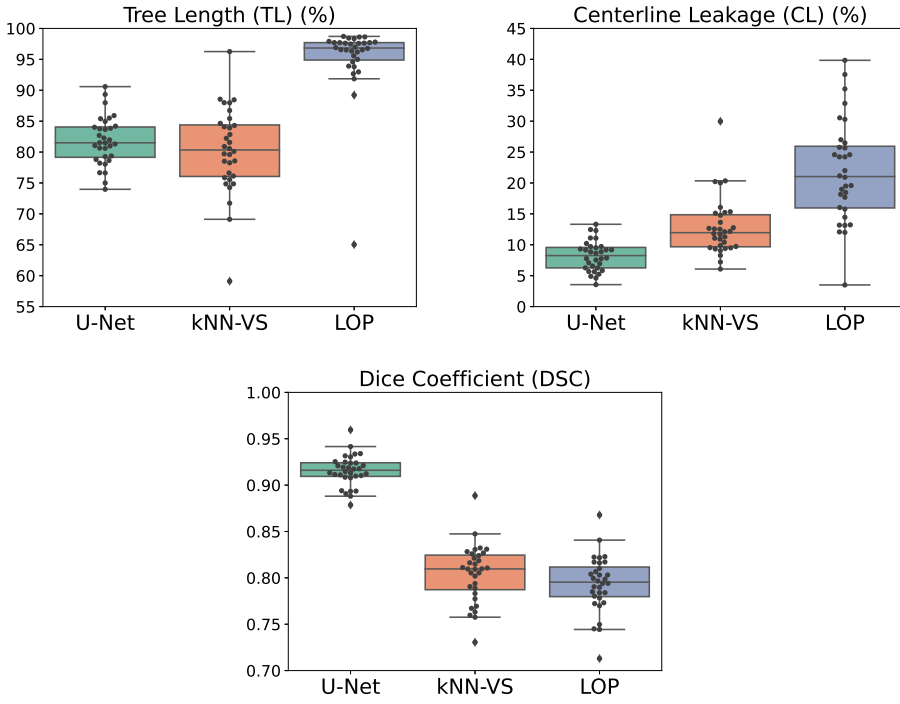


Figure 2.4: Boxplots showing 1) tree length detected, 2) centerline leakage and 3) Dice similarity coefficient on the DLCST data, for the results obtained by our U-Net-based method and the kNN-VS [32] and LOP [33] methods. For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show all the data points.

as possible airways. We found 41 such cases of false positives that were actual airways when they were overlaid and analyzed on the CT scan. Out of these 41 misclassified branches, 20 were segmented only by our U-Net-based method, 14 were segmented by our U-Net-based method and either the kNN-VS or LOP method, and 7 were segmented by either the kNN-VS or LOP method and not the U-Net-based method. All cases segmented by our U-Net-based method were free of leakage. In contrast, all cases segmented by the kNN-VS and LOP methods had errors: they were branches longer than the real airway, or had leakage outside the airway wall. We show in Figure 2.6 some examples of these misclassified false positive branches, where in the corresponding overlay of the airway mask with the CT scan we can see that they are real airways. Because of this, the CL leakage reported for the three methods in Figure 2.4 are presumably overestimated, and to a larger extent for our U-Net-based method. Interestingly, most of the 34 misclassified but correctly segmented branches by our U-Net-based method were on CT scans from subjects with COPD, 26 out of 34 (76%).

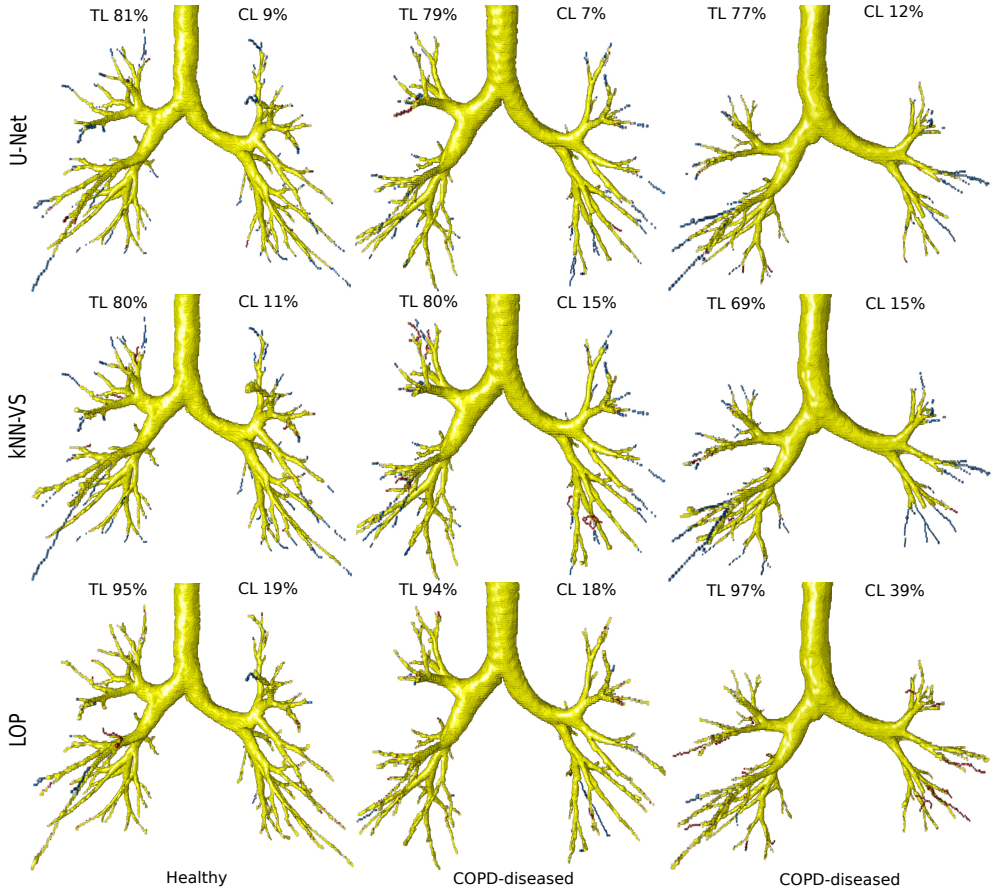


Figure 2.5: Visualization of predicted airways trees on the DLCST data, obtained by our U-Net-based method and the kNN-VS [32] and LOP [33] methods. True positives are displayed in yellow, centerline false negatives in blue and centerline false positives in red. These false negatives and positives are the errors counted in the tree length detected and centerline leakage measures reported in Figure 2.4, respectively. The cases displayed correspond to, from left to right, the two U-Net results with Dice score closest to the median of the test Dice measures, and the U-Net result with the lowest Dice score.

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

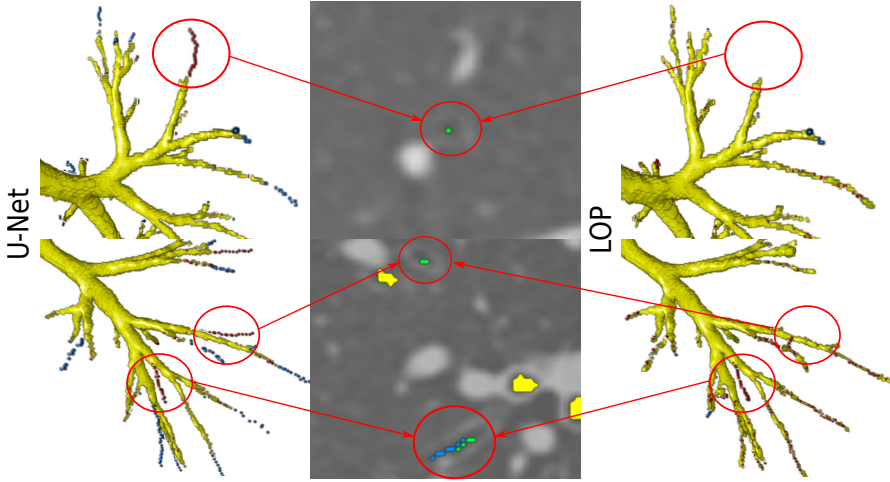


Figure 2.6: Visualization of examples of predicted false positive branches on the DLCST data, obtained by our U-Net-based method and missed by the kNN-VS and LOP methods, that are actually real airways missing in the reference segmentations. The last earmarked branch is detected by the LOP method but with errors, with leakage outside the airway wall. In the 3D visualizations, the true positives are displayed in yellow, centerline false negatives in blue and centerline false positives in red. We also show the overlay of the predicted airway centerline masks with the CT scan, showing that the earmarked branches are real airways. In this overlay, the predicted centerlines by our U-Net method are displayed in green, those by the LOP method in blue, and the reference segmentations in yellow.

measures in Figure 2.B.1b in the Appendix.

2.5.3 Evaluation on EXACT'09 data

We show in Figure 2.7 the evaluation on the EXACT'09 data of our U-Net-based method, the 15 methods from the challenge EXACT'09 [38], the 6 post-challenge methods [85], the 4 recent methods [43, 44, 52, 86] evaluated on the EXACT'09 data, and our experiment with the nnU-Net method [87] on these data. The reported results are the average TL and FPR measures obtained over the EXACT'09 test set. Our U-Net-based method achieves an overall TL of 70.3% and FPR of 2.74%. When compared to the EXACT'09 methods and the 6 post-challenge methods [85], our U-Net-based method has a TL much higher (at least 10%) than the scores from all methods except two, while the reported FPR is only slightly higher ($\approx 1\%$) than the average score among these methods. This indicates that our U-Net-based method predicts more complete airway trees on average, and with limited false positive errors. The two methods with higher TL also show a much higher FPR. When compared to the CNN-Leak method [43] and the nnU-Net method [87], our U-Net-based method has higher completeness but also more false positive errors. The nnU-Net method [87] achieved a similar performance to the CNN-Leak method [43], with slightly lower

TL for a similar FPR score. When compared to the U-Net-FRAD method [52], our U-Net-based method has slightly lower completeness but also less false positive errors. Our method together with the CNN-Leak [43], U-Net-FRAD [52] and nnU-Net [87] methods seem to have the best trade-off between the TL and FPR scores among all tested methods in Figure 2.7. However, the authors from [43] did not follow the independent evaluation for the EXACT'09 benchmark, but they re-implemented the evaluation algorithm using the reference standard from EXACT'09 [38]. Their evaluation on one of the 15 submissions to EXACT'09 resulted in a slightly higher tree length than the one originally reported in [38]. Thus, the TL score reported for this method could be overestimated. Comparing our U-Net with the nnU-Net results, the former shows higher TL (68.8 (61.2–79.7) vs. 63.9 (52.8–75.2), $p=0.01$) and higher FPR (1.90 (0.51–3.86) vs. 1.07 (0.30–2.25), $p=0.04$), but they are not significantly different. This indicates that both our U-Net-based method and the nnU-Net have comparable accuracy. When compared to the 2.5D CNN method [44] and the PICASSO method [86], our U-Net-based method shows better performance measures, with higher completeness and less false positive errors.

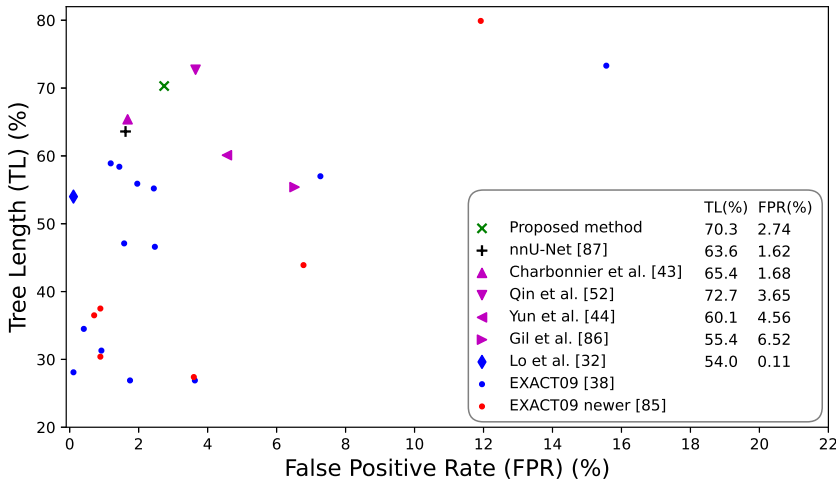


Figure 2.7: Average tree length detected versus average false positive rate over the EXACT'09 test set, for our U-Net-based method, the 15 methods from the challenge EXACT'09 [38], the 6 post-challenge methods [85], the 4 recent methods [43, 44, 52, 86] evaluated on the EXACT'09 data, and our experiment with the nnU-Net method [87] on these data. Results are computed by the EXACT'09 challenge organizers except for [43], who did the evaluation slightly differently, leading to slightly better performance, as reported in their paper.

Through visual inspection of the results, we observed that for some CT scans a large number of airways predicted by our U-Net-based method that were reported as false positives in the EXACT'09 evaluation are real airways on the CT scan. We show in Figure 2.8 visualizations of airway trees, obtained by our method and the nnU-Net

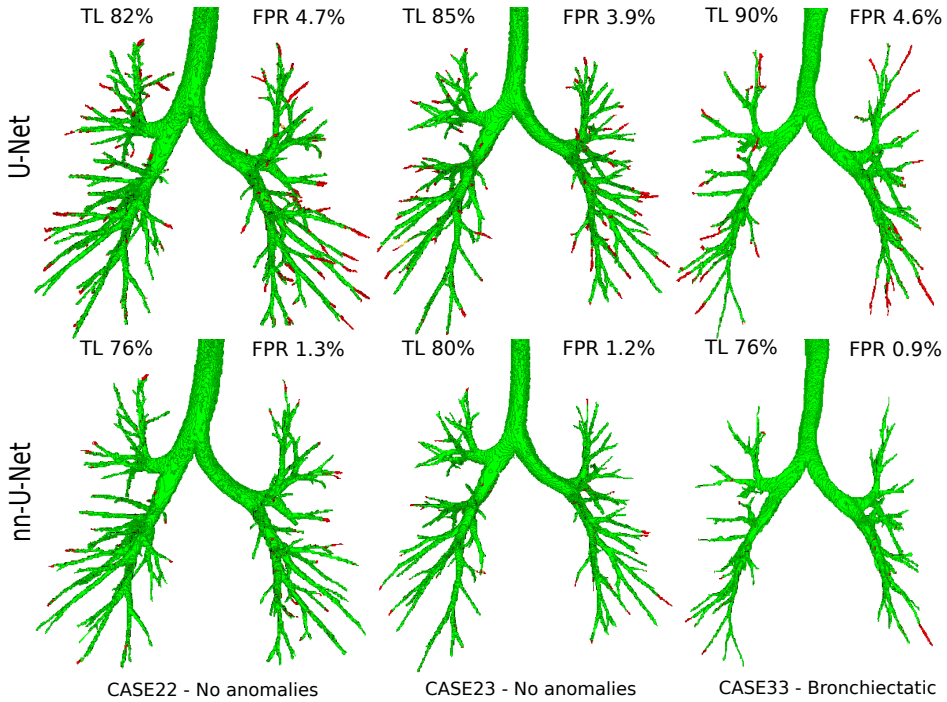


Figure 2.8: Visualization of predicted airway trees on the EXACT'09 data obtained by our U-Net-based method and the nnU-Net method [87]. True positives are displayed in green and false positives in red, indicated according to the EXACT'09 evaluation. The cases displayed correspond to 3 out of the 5 U-Net results with the highest false positive rate reported in the EXACT'09 evaluation, where we found through visual inspection that many of the reported false positives are real airways on the CT scan, which were presumably missing in the EXACT'09 reference.

method [87], for which most of the reported false positive errors are real airways. Our method shows more misclassified false positive branches than the nnU-Net method [87], which partly explain the higher FPR scores in the displayed cases. This is because the reference standard from EXACT'09, constructed from all 15 participating methods to the challenge, is somewhat incomplete. It was estimated that on average approximately 10% more branches were visible in the CT scans than those that were included in the reference [38]. Thus, the false positive rate reported by our U-Net-based method and the nnU-Net method [87] in Figure 2.7 may be overestimated, with a larger extent for our method. This could be also for the other methods [43, 44, 52, 85, 86] we compared with, which used the same reference for evaluation.

We compared the measures from our U-Net-based method between a group of CT scans without any reported anomalies on the CT scan (cases 21, 22, 23, 28, 29, 35, and 37), and a group of CT scans with reported bronchiectasis (cases 33, 34, 36, and 39) [38]. We found no significant differences in TL (80.1 (65.6–83.2) vs. 79.4 (75.7–82.1), $p=0.59$), in FPR (2.98 (0.99–3.87) vs. 3.34 (1.99–6.50), $p=0.21$) and

in total tree length (116 (100–193) vs. 275 (232–295), $p=0.07$). This indicates that our method has similar accuracy with respect to the reference segmentations on both CT scans without anomalies and CT scans with bronchiectasis. We show these grouped measures in Figure 2.B.1c in the Appendix.

2

2.6 Discussion

2.6.1 Performance compared to other methods

The proposed U-Net-based method has shown good performance on the three different datasets, achieving highly complete airway segmentations with low false positive errors. The three datasets tested have very different characteristics: from subjects of a wide age range (from children in CF-CT data to older adults in DLCST data) and with airway abnormalities from diverse lung diseases, including CF and COPD. On the CF-CT and DLCST data, our U-Net-based method obtained more accurate airway segmentations than the other tested methods [32, 33, 62], shown by higher completeness and Dice scores altogether, and lower false positive errors. On the EXACT’09 data, our U-Net-based method achieved performance measures similar to the best performing previous methods, with the second highest sensitivity score among all methods that reported good specificity. Especially, it is noticeable that our method achieved this good performance on the highly heterogeneous EXACT’09 data with models trained on different and more homogeneous datasets (CF-CT and DLCST). This shows the robustness and the capacity to generalize across different data of the proposed method. Furthermore, our U-Net-based method obtained similar performance measures on CT scans from both healthy and diseased subjects, on the CF-CT and EXACT’09 data. On the DLCST data, our method had slightly more false positives errors on CT scans from subjects with COPD, where airway detection can be more challenging due to emphysema. However, on CT scans with COPD we reported a higher number of false positives in the U-Net results that were real airways on the CT scan, which partly explains the lower specificity in the diseased cases. On the CF-CT data, the method segmented more airway branches on CT scans from subjects with CF, probably due to the widening of peripheral airways due to CF-bronchiectasis, which makes them easier to detect on the CT scan.

When compared to other CNN-based methods evaluated on the EXACT’09 data, our U-Net-based method shows a performance similar to that of the CNN-Leak [43], U-Net-FRAD [52] and nnU-Net [87] methods. These four methods seem to have the best trade-off between sensitivity and specificity among all tested methods in Figure 2.7. However, the U-Net-FRAD method [52] was trained on data that included the EXACT’09 training set, using their own manually drawn reference segmentations on these data. Since the CT scans from the EXACT’09 training and test sets have similar characteristics and scanning parameters [38], this gives an advantage to the U-Net-FRAD method [52] over our U-Net-based method, which was trained on different data. On the other hand, the CNN-Leak method [43] requires a coarse airway segmentation as initialization to the CNN-based leakage removal algorithm, and thus the completeness of the results is limited by that of the initial segmentation. In contrast, our U-Net-based method segments the full airway tree directly. Moreover, CNN-Leak [43] applied

leakage removal to a series of 15 runs of the region growing based algorithm [90], with different parameters that control the extent of leakage, upon which the results were merged. In fact, the results from applying the leakage removal only to the baseline segmentation [90] had a much lower sensitivity (TL 51.8% in contrast to 65.4%) and a slightly higher specificity (FPR 1.01% in contrast to 1.68%). Similarly, an ensemble of U-Net results with different settings would likely have a slightly better performance, but this is more time consuming. When compared with the nnU-Net method [87], our U-Net-based method achieved slightly higher completeness and false positive errors, but we found that the differences were not statistically significant. When compared with the 2.5D CNN method [44], our U-Net-based method shows better performance with both higher sensitivity and specificity. The higher accuracy of our method can be because the 2.5D CNN method processes three perpendicular 2D slices around each voxel, while the 3D U-Net can better capture the 3D morphological information of airways.

Regarding computational efficiency, our U-Net-based method has a low execution time during inference of about 1 min per scan, including all post-processing steps to obtain a single connected binary airway tree. It should be noted that computation times of different methods cannot easily be compared due to differences in hardware, but the following may give a rough idea of the efficiency of our method compared to previous work. The execution times reported by other methods evaluated on the EXACT'09 data are: the kNN-VS method [32], 55 min per scan, although it was run on a much older CPU and was not parallelized; the CNN-Leak method [43], 3–5 min per scan; the U-Net-FRAD method [52], approximately 50 s per scan, although it does not include any post-processing steps; the 2.5D CNN method [44], 2–8 min per scan; and the PICASSO method [86], approximately 10 min per scan. Our experiments with the nnU-Net method [87] required 5–15 min per scan, although this includes some expensive pre-processing operations applied on the fly to the test images, such as resampling. Additionally, the U-Net method in [55] reports approximately 6.5 min per scan.

2.6.2 Advantages of the proposed method

The proposed U-Net-based method processes large 3D patches extracted from the CT scan in a single pass through the network. This makes our method simple, robust, and efficient at inference time, as only a few large patches are processed to segment a full CT scan. Although processing large input patches does not necessarily imply better performance, we observed that using large input patches resulted in better results and faster convergence of the training and validation losses. This is likely due to the more efficient use of the data during training. In contrast, when using smaller patches to train the model, this requires sampling patches at every voxel to equally process the same region of the CT scan comprised in a larger patch, which results in a high number of total patches. Moreover, since convolution operations are accelerated on the GPU, it is more efficient computationally to process a large patch at once through the network, rather than sequentially loading to GPU and processing smaller patches to include the same large region of the CT. In contrast to our approach, other U-Net-based methods in the literature either 1) apply a small U-Net locally around detected candidate

airway locations, processing many image patches per CT scan [45, 46], or 2) apply a model with larger memory footprint on smaller image patches in a sliding-window fashion [47, 48, 49, 50, 51, 52, 53, 54, 55]. With our simple U-Net-based method and without further processing of the network output (except for computing the largest connected component) we obtained highly complete and accurate airway segmentations on the three datasets tested. On the EXACT’09 data, our method achieved a similar performance to the more complex U-Net-FRAD method [52], which used manually annotated EXACT’09 data for model training, and the nnU-Net method [87], using a similar U-Net as backbone. An additional advantage of processing few, large patches by our method is that fewer edge artifacts are introduced when tiling together the predicted output patches of the network. These artifacts typically occur where the tiled predicted patches meet in the full-size output (or where the patches overlap if this occurs), because the predictions from each patch can be slightly different due to border effects when using non-valid convolutions, and can cause discontinuities in the predicted airway mask.

2.6.3 Limitations

A limitation of our validation of the proposed method is that we did not evaluate it on CT scans with severe airway disease. The CF-CT data include subjects with moderate CF disease, the DLCST data include subjects with moderate COPD, and the EXACT’09 test data have various airways abnormalities but only one CT scan with reported “extensive bronchiectasis” [38]. Testing the method on severe cases would be important to assess its generalizability to tackle challenging airways with abnormally deformed shapes due to severe disease. Moreover, our evaluation of segmentation performance with respect to the presence of disease on the DLCST dataset may not show the whole picture. This is because the reference segmentations on these data were built in a conservative way from automatic airway extractions [32, 33], and could be less complete for CT scans with severe emphysema.

A limitation of our U-Net-based method is that the prediction of the airway tree output by the U-Net is not guaranteed to form a connected tree structure. This could complicate the automatic extraction of airway biomarkers based on these segmentations as some methods assume a fully connected tree as input [61, 62, 65]. The airway predictions obtained in this work had typically some segmented peripheral airways disconnected from the main airway tree, and these were discarded when computing the largest connected component, which reduced the completeness of our airway segmentations. Alternatively, airway measurements techniques that do not rely on fully connected trees, such as [91], could be used. Several methods have been proposed to extract more complete, connected tree structures. The voxel-connectivity U-Net formulation [47] aims to improve connectivity in the airway prediction, resulting however in a model that is significantly more complex than our U-Net used. The linear programming-based tracking module on top of U-Net [49] attempts to link disconnected components of airways from the U-Net output. The mean-field networks and graph neural networks [80] emphasize the prediction of connected tree-like structures by phrasing the tree extraction problem as graph refinement starting from an over-connected input graph. The joint U-Net-GNN method [42] attempts to integrate

this tree-like modeling in the U-Net prediction. However, none of these methods can guarantee fully connected airway tree predictions.

Finally, the memory footprint of our U-Net-based method could be further reduced, which would allow us to fit even larger images to the network, and possibly the entire CT scan. It may be possible to reduce the number of feature maps, especially in the decoding path of the U-Net, without decreasing much the performance. Also, using the partially reversible U-Net formulation [92] in our method could largely reduce its memory footprint, by lowering the number of activation maps in the network stored in memory. However, this may result in an increase of training time (the authors from [92] reported a 50% increase for their tested 5-level U-Net model, similar to ours).

2.7 Conclusions

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

Author contributions

A. Garcia-Uceda: project lead, literature research, experiments, results analysis and writing. R. Selvan: contribution to results analysis. Z. Saghir: project data. H. Tiddens: project data. M. de Bruijne: lead project manager, supervision of literature research, experiments, results analysis and writing.

Availability of data and materials

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

Funding

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

Appendix

2.A Learning curves of the proposed method

2

We computed the learning curve for the proposed U-Net-based method trained on both CF-CT and DLCST data together. To do this, we trained several models with different sizes of the training data. The maximum training size has half of the CT scans from the CF-CT and DLCST datasets (28 scans in total), which is the same data we used to train the model evaluated on the EXACT'09. We keep the remaining 28 CT scans for testing the trained models. For each training set used, the ratio between CF-CT and DLCST CT scans is the same as in the full dataset. We did three experiments for each training size, with randomly assigned training images (except for the largest run with 28 CT scans). To compute the airway predictions on the test data, we did not extract the largest connected component from the thresholded output of the U-Net, as we did for the other experiments. This is to account for the full prediction of the U-Net in assessing the method accuracy for all training sizes. To compare the results for different training sizes, we applied the paired, two-sided Student's T-test on the average of the measures from the three experiments for a given size, and consider that a p -value lower than 0.01 indicates a significant difference. We show in Figure 2.A.1 the computed learning curves, with the different performance metrics obtained for each run and training set size. The measures of tree length detected increase progressively with the training size. The difference between the scores with sizes of 18 and 28 images is still significant ($p < 0.001$), and adding more training images could still improve slightly the results. For the measures of centerline leakage and Dice coefficient, they are more similar between sizes of 9 and 18 images ($p = 0.35$ and $p = 0.26$, respectively) and between sizes of 18 and 28 images ($p = 0.99$ and $p = 0.019$, respectively).

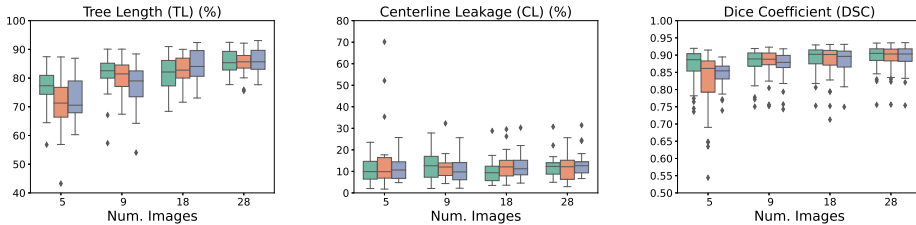
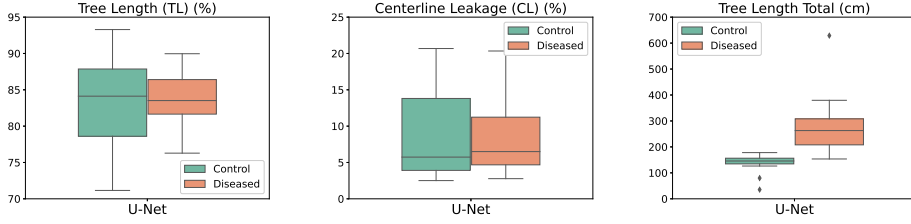
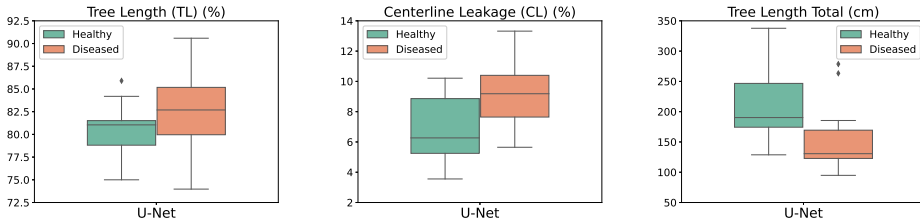


Figure 2.A.1: Learning curve for our U-Net-based method trained on both CF-CT and DLCST data together. Boxplots showing 1) tree length detected, 2) centerline leakage and 3) Dice similarity coefficient on both CF-CT and DLCST data together, for each experiment and training size. For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show the data outliers.

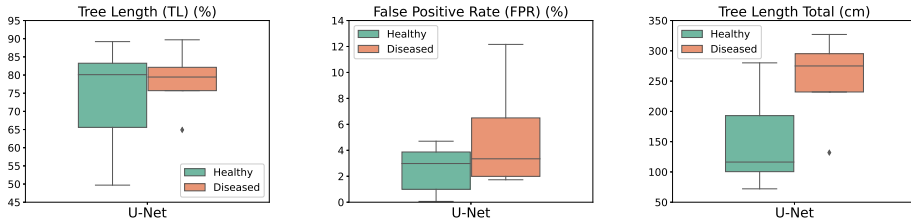
2.B Results grouped by the presence of lung disease



(a) on CF-CT data, grouped by CF disease.



(b) on DLCST data, grouped by COPD disease.



(c) on EXACT'09 data, grouped by bronchiectasis.

Figure 2.B.1: Boxplots showing 1) tree length detected, 2) centerline leakage or false positive rate, and 3) total tree length detected, grouped by the presence of lung disease in the CT scans from each dataset, for the results obtained with our U-Net-based method. For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show the data outliers.

2.C Implementation details of the nnU-Net method

The nnU-Net method proposed by Isensee et al. [87] (<https://github.com/MIC-DKFZ/nnUNet>) is a general segmentation framework designed for biomedical segmentation tasks. We applied this method for airway segmentation from chest CT scans.

For our experiment, we used the so-called “full3d_UNet” in the nnU-Net framework, which is the most similar to the U-Net in our method. This U-Net has 5 levels of resolution, where each level in the downsampling / upsampling path has two $3 \times 3 \times 3$ convolutional layers and a $2 \times 2 \times 2$ pooling or deconvolution layer, respectively. Each convolution operation is followed by an instance normalization layer and a leaky rectified linear (leaky-ReLU) activation. The number of features in the top resolution level is 32, and after each pooling or deconvolution layer the number of feature channels is doubled or halved, respectively. The network uses deep supervision at all levels of the U-Net. In this, the output of every last convolutional layer at every resolution level is concatenated, after being resampled to the original resolution. Then, the final layer is a $1 \times 1 \times 1$ convolutional layer followed by a sigmoid activation function.

To train the network, we used the same 28 CT scans and ground truth segmentations from the CF-CT and DLCST datasets as we used to train our method. The nnU-Net method uses as training loss function a combination of the binary cross entropy and soft Dice losses. We could not modify the loss computation in the nnU-Net to consider only voxels within the lung regions, as we did for our method in Equation 2.1. Instead, we masked the ground truth segmentations to the mask of the lung fields, to remove the trachea and part of the main bronchi. The nnU-Net method uses the SGD optimizer with an adaptable learning rate, starting with a value of 1×10^{-2} . We trained the model for a sufficiently large number of epochs, 600, until the training and validation losses are clearly stabilized. We then retrieved the model with the overall minimum validation loss for testing, denoted by “model_best” in the nnU-Net framework. Training time was approximately 2–3 days on a GPU GeForce RTX 2080 Ti. Test time inference takes between 5–15 min per scan, including pre-processing.

The nnU-Net applies some pre-processing operations to the CT scans and ground truth in the training dataset. First, the images are cropped to the region of non-zero values in the ground truth airway masks. Then, the images are resampled to a fixed resolution equal to the median over the training dataset, using 3rd order spline interpolation for the CT scans and nearest neighbor interpolation for the ground truth masks. During training, the nnU-Net extracts random patches from the CT scans and ground truth, with a total of 1 patch per image and per epoch. Then, random rigid transformations are applied for data augmentation, including 1) flipping in the three directions, 2) random small 3D rotations, and 3) random scaling.

At inference time, the nnU-Net applies the same pre-processing operations to the test CT scans as those applied on the training data used for the tested model. The images are resampled to the same fixed resolution used for the training data, i.e., the median over these data. Input patches to the network are extracted from the CT scans in a sliding-window fashion, with an overlap of roughly 50% between the patches in each direction, and then are processed through the trained model. The output predicted patches containing airway probability maps are aggregated and the full-size output is reconstructed. Then, thresholding is used to obtain the airway binary segmentation, and the pre-processing steps are reversed to recover the original image resolution. Finally, we merge this segmentation with a mask for the trachea, main bronchi, and the first 5 voxels of the next branches to obtain the full airway tree, which is easily computed by a region growing method [32].



CHAPTER 3

Label refinement network from
synthetic error augmentation for
medical image segmentation



Abstract

Deep convolutional neural networks for image segmentation do not learn the label structure explicitly and may produce segmentations with an incorrect structure, e.g., with disconnected cylindrical structures in the segmentation of tree-like structures such as airways or blood vessels. This chapter presents a novel label refinement method to correct such errors from an initial segmentation, implicitly incorporating information about label structure. This method features two novel parts: 1) a model that generates synthetic structural errors, and 2) a label appearance simulation network that produces synthetic segmentations (with errors) that are similar in appearance to the real initial segmentations. Using these synthetic segmentations and the original images, the label refinement network is trained to correct errors and improve the initial segmentations. The proposed method is validated on two segmentation tasks: airway segmentation from chest computed tomography (CT) scans and brain vessel segmentation from 3D CT angiography (CTA) images of the brain. In both applications, our method significantly outperformed a standard 3D U-Net and other previous refinement approaches. Improvements are even larger when additional unlabeled data is used for model training. In an ablation study, we demonstrate the value of the different components of the proposed method.

Based on: S. Chen*, **A. Garcia-Uceda***, J. Su*, G. van Tulder, L. Wolff, T. van Walsum, and M. de Bruijne, “Label refinement network from synthetic error augmentation for medical image segmentation,” *Submitted*

* indicates equal contributions

3.1 Introduction

Convolutional neural networks (CNNs) are the state-of-the-art for many biomedical imaging segmentation tasks. Many CNN segmentation architectures have been proposed, such as fully connected networks [76], Dense-Net [93] and the U-Net [56]. The U-Net has become the most popular network for biomedical image segmentation, due to its efficient structural design featuring skip-connections, showing superior accuracy and robustness in various segmentation tasks [87, 94]. Most CNN-based segmentation methods including the U-Net do not fully exploit and encode the structural information of the objects to be segmented. Consequently, these methods may produce segmentations with errors that become obvious when looking at the full segmented structure. Examples of such errors are discontinuities in the segmentations of elongated tubular structures, such as airways in the lungs, as shown in Figure 3.1. Using label structural knowledge such as continuity in the branches of the airway tree can help prevent these errors. However, it is not trivial to explicitly encode this global information in CNNs.

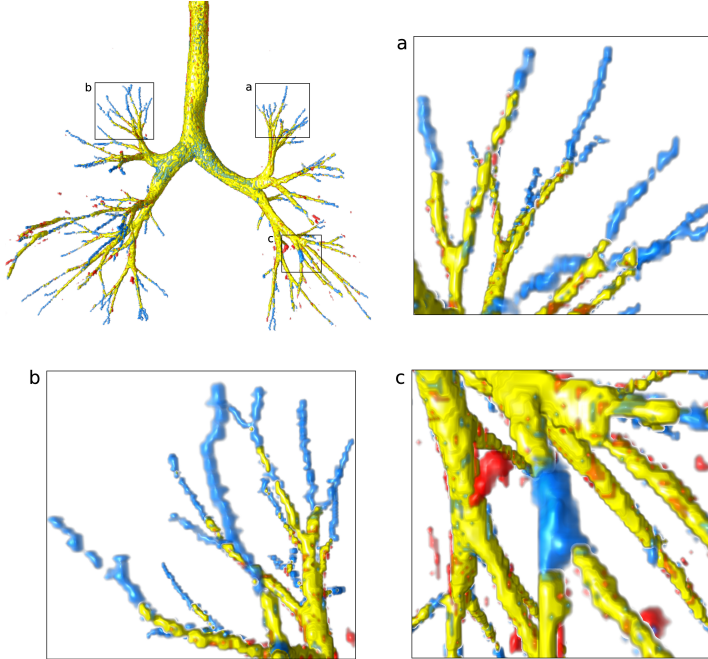


Figure 3.1: Common structural errors in the segmentations obtained by a U-Net, trained to segment airways in the lungs [41]. True positives are displayed in yellow, false negatives in blue and false positives in red. Detailed views a-b show errors as missing terminal branches, and view c shows a discontinuity error in the branch.

In this work, we propose a framework to implicitly encode the label structural information into CNNs by formulating this as a label refinement step. Specifically, we generate structural errors in labels (such as the ground truth or initial segmentations)

and train a label refinement network to correct these errors. The trained network is expected to generalize to the real errors in the initial segmentations produced by a baseline segmentation network and correct them. To enhance the generalizability of the label refinement network on the initial segmentations, a label appearance simulation network is applied to reduce the appearance difference between the synthetic labels and the initial segmentations. With these synthetic labels (and the initial segmentations) together with the original image as inputs and the ground truth segmentations as reference, the label refinement network can learn to correct those errors and incorporate this in its segmentation decisions.

We validated the proposed label refinement method on two segmentation tasks: airway segmentation from chest computed tomography (CT) scans [41] and brain vessel segmentation from 3D CT angiography (CTA) images of the brain [95]. We compared our method with a U-Net baseline and other refinement networks, including DoubleU-Net [96] and SCAN [97], an adversarial refinement network. Moreover, we conducted an ablation study to show the contribution of each individual component of the label refinement method. Finally, we performed experiments in a semi-supervised setting to train our method using additional unlabeled data.

3.1.1 Related work

Label refinement

In this work, we apply a refinement network on the initial segmentation from a baseline segmentation network together with the original image, with the aim of correcting errors in the initial segmentation. A similar approach has been used in other previous works. Jha et al. [96] attached a second U-Net network to a baseline U-Net, using as inputs the original image multiplied with the output of the first U-Net. Yang et al. [98] refined low-quality manual annotations made by non-experts by training their method with added noise in order to reduce the inter-observer inconsistency of the annotations. Unlike our method, Yang et al. do not focus on refining an initial automatic segmentation and therefore the label appearance simulation network is not needed. Dai et al. [97] refined the segmentations from a fully convolutional network by using adversarial training to reduce the domain gap between the target predictions and the ground truth segmentations on training data. Araujo et al. [99] attached a variational auto-encoder after a U-Net network to encode the label topology of the ground truth segmentations for a better label reconstruction. Different from Dai et al. and Araujo et al., our work does not focus on encoding [99] or discriminating [97] the overall label topology, but instead on learning to correct the most common errors in the segmentations.

Airway segmentation

The airway tree in the lungs forms a complex 3D tree-like branching network, with many branches of different sizes and orientations. The peripheral branches of smaller size are challenging to segment from chest CT scans, as they have obscured borders due to partial volume effects. Many classical methods for airway tree extraction are based on a region growing algorithm [30, 32, 33]. However, their accuracy is limited,

and they typically miss a large number of the smaller peripheral airways [38]. Many state-of-the-art airway segmentation methods are based on CNNs, and especially the U-Net [41, 52, 100, 101]. CNN-based methods can obtain more accurate and complete segmentations than previous intensity-based methods. However, even the latest U-Net-based methods usually miss several terminal branches, and make errors in continuity around the smaller segmented branches.

Brain vessel segmentation

The brain vessels form a complex 3D branching network that consists of veins and arteries. In 3D CTA images of the brain, many seemingly isolated vessel structures can be present due to the image acquisition and vascular diseases, such as ischemic large vessel occlusions. State-of-the-art vessel segmentation methods have been applied to 3D time-of-flight (TOF) magnetic resonance angiography (MRA) images [102, 103, 104], and to 3D and 4D CTA images [105] using U-Nets. Su et al. [95] used a U-Net-based method to extract a dilated vessel centerline approximation. Compared to previous vessel segmentation methods [102, 103, 104, 105], centerline extraction recovers the topology of the vessel structure more accurately (e.g., “kissing vessels” appear connected in the full segmentations but are disconnected in centerline extraction). However, the U-Net still makes other topological errors such as local connectivity gaps in vessel branches.

3.2 Methods

3.2.1 Overview

The proposed method consists of four steps, schematically shown in Figure 3.2. Firstly, a baseline segmentation network generates the initial segmentations (Section 3.2.2). Secondly, synthetic errors are generated and added to every ground truth segmentation, in order to generate synthetic labels to train the label refinement network (Section 3.2.3). Thirdly, a label appearance simulation network (LASN) based on adversarial learning is used to reduce the appearance difference between the synthetic labels and the initial segmentations (Section 3.2.4). Finally, a label refinement network is trained to predict the final segmentation, using the synthetic labels (or the initial segmentations) and the original images as inputs, and the ground truth segmentations as reference (Section 3.2.5).

3.2.2 Base segmentation network

We use a base segmentation network f_1 to predict an initial segmentation. Given a medical imaging dataset that contains an image I and the ground truth segmentation g for each subject, the model $f_1(I|\theta_1)$, with θ_1 the trainable parameters, is trained by minimizing the Dice loss $\mathcal{L}_1 = \mathcal{L}_{dc}(f_1(I), g)$,

$$\mathcal{L}_{dc}(y, g) = -\frac{2 \sum_{i \in I} y_i g_i}{\sum_{i \in I} y_i + \sum_{i \in I} g_i} \quad 3.1$$

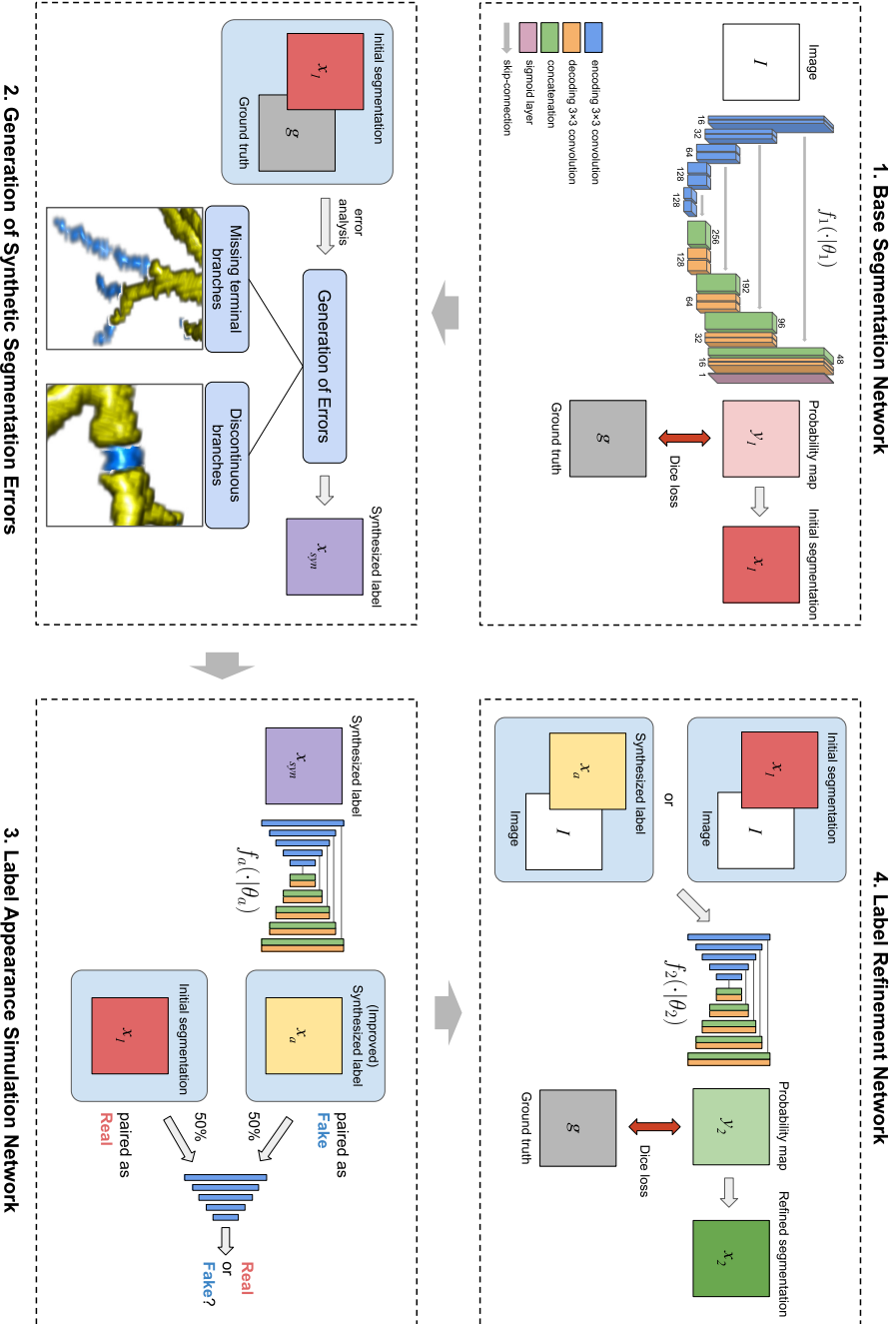


Figure 3.2: Schematics of the proposed label refinement method. First, a base segmentation network f_1 is trained to obtain the initial predictions x_1 . Second, we synthesize a new dataset x_n that contains similar errors to x_1 . Third, a label appearance simulation network f_a (together with a discriminator D) is trained to obtain a more realistic dataset x_a . Finally, the label refinement network f_2 is trained with x_1 and x_a together with the image I as inputs.

where y_i and g_i are the i th voxel values of the probability maps output by the model (in this case $y = f_1(I)$), and the ground truth segmentation, respectively.

The initial predicted segmentation is x_1 , obtained by thresholding the output probability maps y_1 of the network with value 0.5. x_1 may contain label structural errors, such as discontinuous branches in a tree-like structure. Next, we show how to design synthetic errors similar to those in x_1 that can be used to train the label refinement network.

3.2.3 Generation of synthetic segmentation errors

We use synthetic labels \mathbf{x}_{syn} with added synthetic errors to train the label refinement network. Depending on the experimental set-up, the errors can be added to the ground truth or to the initial segmentations. The synthetic errors are generated to resemble those in the initial segmentations x_1 , based on our initial analysis of common errors. In this work, we focus on two structures: airways in the lungs and vessels in the brain. Airways and vessels share several characteristics: they both form 3D branching networks, with branches of cylindrical shape and various sizes and orientations. We use this prior shape knowledge to generate synthetic errors, as described below for each structure.

Synthetic errors for airways

Most of the errors in airway segmentations can be grouped into two types: 1) missing terminal branches, partially or totally, and 2) discontinuity in the segmented branches, which occurs more frequently in smaller branches. Examples of errors in airway segmentations obtained by the baseline segmentation network in Section 3.2.2 are shown in Figure 3.1. To generate similar, synthetic errors, we select a random subset of branches in the airway tree and partially remove the segmentation of the selected branches by masking it at a random position and with a random length. Branches are identified using the airway centerline tree, extracted from the airway segmentation [38]. Single branches are defined as the segments between two bifurcation points or between the last bifurcation and the end of terminal branches. The applied masking is defined differently for each type of error:

Missing terminal branches: The subset of branches in which to synthesize errors is randomly sampled from all the terminal branches in the airway tree, defined as branches with no further bifurcations downstream. A mask of cylindrical shape is applied to (partially) remove the selected branch. The mask is defined by 1) a start point, that is a random position along the branch centerline between the branch start and middle points; 2) a length, that is the distance between the mask start point and branch end; and 3) a width, that is three times the branch diameter.

Discontinuity in branches: The subset of branches with errors is randomly sampled from all the branches in the airway tree, excluding the trachea, the two main bronchi and the 2nd generation airways, and including the terminal branches. We assign a higher sampling probability to branches of higher airway generation, where the generation is defined as the number of branch bifurcations counted in the path linking the given branch and the root of the airway tree, i.e., the trachea. The sampling

probability p_i for each candidate branch is defined as $p_i = g_i / \sum_{k=1}^{N_c} g_k, \forall i = 1 \dots N_c$, where g_i is the airway generation and N_c the number of candidate branches. A mask of cylindrical shape is applied to create a gap in the selected branch. The mask is defined by 1) a center, that is a random position along the branch centerline; 2) a length, that is a random distance between a minimum of 10 voxels and the total branch length; and 3) a width, that is three times the branch diameter.

Parameters: The extent of each type of errors in the airway synthetic labels is determined by a separate parameter, denoted as p_1^a and p_2^a . p_1^a is the proportion of selected branches with errors of type “missing terminal branches”, with respect to all the terminal branches. p_2^a is the proportion of selected branches with errors of type “discontinuity in branches”, with respect to all the branches in the airway tree.

Synthetic errors for brain vessels

Most of the errors in brain vessel segmentation are in the form of incomplete or missing vessel branches. To generate similar, synthetic errors, we create random discontinuous gaps in the segmentation of each vessel by masking it at a random position and with a random length. Since the errors occur more frequently for long vessels than for short ones, we group all the vessels into three equal-sized groups: long, medium size and short, based on the relative centerline segment lengths in each subject. The distribution of vessel lengths (in voxels), using the median and interquartile range (IQR), is: for long segments 70 (49–106), for medium-size segments 29 (22–36), and for short segments 13 (9–17). For long segments, the maximum number of injected gaps is 6 (randomly sampled from a uniform distribution between 0 and 6 positions) with gap length between 10–35 voxels. For medium-size segments, the maximum number of gaps is 4 with gap length between 10–20 voxels. For the short segments, the maximum number of gaps is 2 with gap length between 6–15 voxels. Those error injections are applied on the 1 voxel-wide ground truth centerlines, by dilating it with a $3 \times 3 \times 3$ cubic structure element to generate the final vessel synthetic label.

Parameters: The extent of errors in the vessel synthetic labels is determined by only one parameter, denoted as p^v . p^v is the proportion of selected branches with errors with respect to all the branches in the vessel network.

3.2.4 Label appearance simulation network

Although the synthetic labels x_{syn} are designed to have similar structural errors to the initial segmentation x_1 , there may be an appearance difference between x_1 and x_{syn} (see an example in Figure 3.3). The label refinement network trained on x_{syn} may therefore generalize poorly to x_1 . To prevent this, we use a label appearance simulation network $f_a(\cdot|\theta_a)$ to change the appearance of x_{syn} to be more similar to that of x_1 , while preserving the synthetic errors that we added to x_{syn} .

The label appearance simulation network $f_a(\cdot|\theta_a)$, with θ_a the trainable parameters, is optimized by adversarial learning via a discriminator D ,

$$f_a^* = \arg \min_{f_a} ((\max_D \mathcal{L}_{\text{adv}}(f_a, D)) + \lambda \mathcal{L}_{\text{dc}}(x_a, x_{\text{syn}})) \quad 3.2$$

with the adversarial loss \mathcal{L}_{adv} defined as,

$$\mathcal{L}_{\text{adv}}(f_a, D) = \mathbb{E}_{x_1}[\log D(x_1)] + \mathbb{E}_{x_a}[\log(1 - D(x_a))] \quad 3.3$$

where D is a classifier, discriminating between the given label x and the initial segmentation x_1 . It outputs a probability between 0.0 and 1.0. $x_a = f_a(x_{\text{syn}})$ is the appearance-enhanced label of x_{syn} . We added a Dice-based identity loss $\mathcal{L}_{\text{dc}}(x_a, x_{\text{syn}})$ to train $f_a(\cdot)$, in order to preserve the synthetic errors that we added in x_{syn} . The hyperparameter λ controls the balance between the identity loss and the dissimilarity adversarial loss.

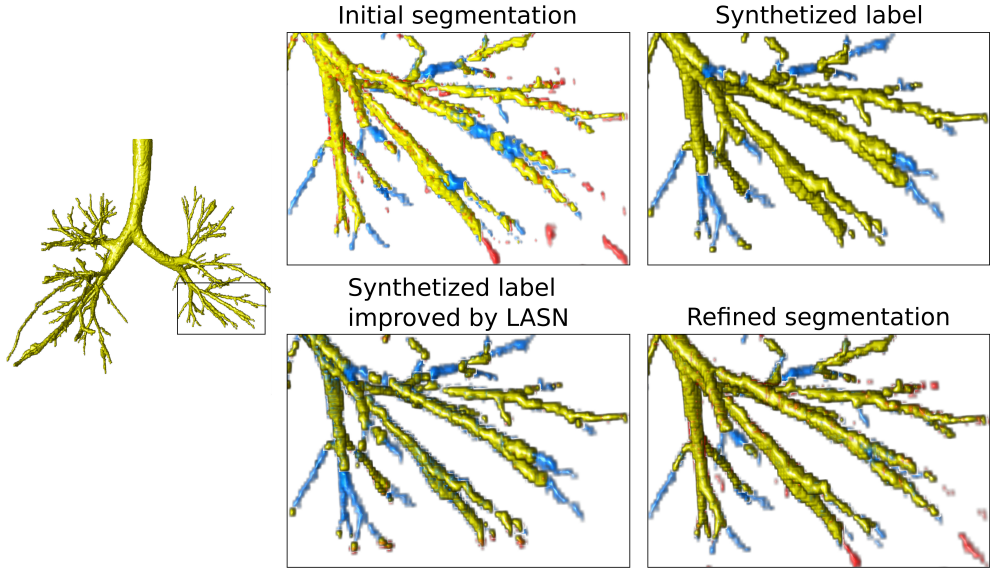


Figure 3.3: Example of segmentation of airways in the lungs obtained by the different components of the proposed method. In the detailed views, true positives are displayed in yellow, false negatives in blue and false positives in red.

3.2.5 Label refinement network

Finally, we optimize a label refinement network f_2 to predict the ground truth segmentations, based on the synthetic labels with errors x_a together with the original image as inputs. This way, f_2 learns to correct segmentation errors and can be used to improve the initial segmentations x_1 . The model $f_2(I|\theta_2)$, with θ_2 the trainable parameters, is trained by minimizing the Dice loss $\mathcal{L}_2 = \mathcal{L}_{\text{dc}}(f_2(I, \tilde{x}), g)$, given by Equation 3.1. The final segmentation result is x_2 , obtained by thresholding the output probability maps y_2 of the refinement network with value 0.5.

3.3 Experiments

3.3.1 Datasets

We validated the proposed method on two biomedical imaging segmentation tasks: segmenting airways from chest CT scans and brain vessels from CTA images of the brain.

Chest CT data

The dataset of chest CT scans is from a retrospective study of pediatric patients (6 to 17 years old) with cystic fibrosis lung disease, acquired routinely at the hospital Erasmus MC-Sophia Rotterdam [106]. The CT scans show noticeable structural airway abnormalities resulting from the disease. In our study, we used 178 low-dose CT scans acquired at full inspiration breath-hold. All CT scans have slice dimensions 512×512 , with a variable number of slices between 200–1000. Each CT scan has an in-plane voxel size in the range 0.35–0.65 mm, with slice thickness between 0.75–1.0 mm, and slice spacing between 0.3–0.8 mm. A random subset of 65 CT scans from the total 178 scans have annotations of the airway lumen. To obtain these annotations, Thirona’s lung quantification software LungQ (Thirona, Nijmegen, the Netherlands) was used to automatically extract the airway lumen from the CT scan. Then, these segmentations were visually checked by trained data analysts for accuracy, and corrected as needed.

For our experiments, we used as testing data 41 random CT scans from the subset of 65 CT scans with ground truth segmentations. From the remaining 24 CT scans with annotations, we used three different random data splits with 20 CT scans for training the networks and 4 CT scans for validation. The remaining 113 CT scans without ground truth segmentations were used as unlabeled training data for the experiments with semi-supervised learning.

CTA data of the brain

The dataset of CTA images of the brain is from the MR CLEAN Registry [107], an ongoing registry for patients who underwent endovascular treatment for acute ischemic stroke in one of 19 hospitals in the Netherlands since March 2014. The data was collected during clinical practice, and we applied the following data inclusion criteria: 1) slice thickness ≤ 1.5 mm, 2) slice spacing ≤ 1.5 mm, 3) the contrast acquisition phase has to be peak arterial phase, equilibrium or early venous phase [108], and 4) the image should cover at least half of the brain. In our study, we used 69 CTA images from 69 different subjects used in [95]. All CTA images were skull-stripped with an atlas-based registration method [109]. 20 CTA images had no vessel annotations, 9 CTA images had a complete brain vessel centerline annotation, and the remaining 40 CTA images (randomly sampled from the whole dataset) had vessel centerline annotations in a randomly sampled sub-volume of $140 \times 140 \times 140$ voxels. The centerline annotations were dilated with a $3 \times 3 \times 3$ cubic structure element to obtain the ground truth segmentations. Each CTA image has an in-plane voxel size in the range 0.4–0.68 mm, with slice thickness between 0.5–1.5 mm, and slice spacing between 0.3–1.0 mm.

For our experiments, we used as testing data 2 random full-volume CTA scans and 20 random CTA cubes from the set of 9 full-volume, annotated CTA scans and 40 CTA cubes, respectively. From the remaining data with annotations, we used three different random data splits with 7 full-volume CTA scans and 14 CTA cubes for training the networks, and 6 CTA cubes for validation. The remaining 20 full-volume CTA scans without manual annotations were used as unlabeled training data for the experiments with semi-supervised learning.

3.3.2 Parameters for generating synthetic errors

The generation of synthetic errors depends on the parameters p_1^a and p_2^a for airways, and p^v for vessels, described in Section 3.2.3. We will refer to these parameters as “synthetic error rate”, for each type of error. For each training sample, the synthetic error rate is randomly sampled from a uniform distribution between 0.0 and the upper bound, or maximum synthetic error rate. These upper bounds are hyperparameters for the proposed method, denoted as P_1^a and P_2^a for airways, and P^v for vessels.

We conducted experiments varying the hyperparameters for the error generation in the proposed method, i.e., the maximum synthetic error rates (P_1^a and P_2^a for airways, and P^v for vessels), to investigate their influence in the method performance. The results are shown in Section 3.4.3 below.

In our further experiments, the optimal hyperparameters were determined on the validation set for each of the three random data splits that we used, for both applications. Each hyperparameter was searched independently, from 0.0 to 1.0, while fixing the parameters for other error types to 0.0.

3.3.3 Network architecture

The baseline segmentation network f_1 is a 3D U-Net [78], shown in Figure 3.2. The label refinement network f_2 and the label appearance simulation network f_a use a similar U-Net layout, with the discriminator D in f_a using the same layout as the U-Net encoder. The U-Net consists of an encoding path followed by a decoding path, with skip-connections linking the two paths. The network has 5 levels of depth, 16 feature channels in the first layer, and an input image size of $128 \times 128 \times 128$. Each level of the encoding / decoding paths consists of two $3 \times 3 \times 3$ convolutional layers followed by a $2 \times 2 \times 2$ pooling or upsampling layer, respectively. Each convolutional layer consists of $3 \times 3 \times 3$ convolution with zero-padding followed by instance normalization and leaky rectified linear (leaky-ReLU) activation. The number of feature channels is doubled or halved after every pooling or upsampling layer, respectively. The last layer of the U-Net is a $1 \times 1 \times 1$ convolution, combining the outputs into a single feature map, followed by a sigmoid activation. A training batch contains only one image due to GPU memory limits. The networks are implemented using PyTorch [82].

3.3.4 Details of training and inference of networks

For training, we first apply random rigid transformations as data augmentation, in the form of 1) random 3D rotations up to 30 degrees for all axes, 2) random scaling

with factor between 0.7–1.4, and 3) random flipping in the three directions. Then, we generate samples by extracting random image patches of size $128 \times 128 \times 128$ on the fly from the input training images and corresponding ground truth segmentations. For the airway segmentation experiments, a lung mask is applied to the output of the network and the ground truth patches before computing the training loss. For this operation, we use a pre-computed lung mask that is easily obtained with a region growing algorithm [32]. During training, we used the Adam optimizer [88] with an initial learning rate of 1×10^{-2} . To train the refinement network f_2 , the label \tilde{x} in each training sample is randomly sampled with equal probability from either the initial segmentation x_1 or the synthetic label x_a after the label appearance simulation network.

During inference on new images, the input patches are extracted in a sliding-window fashion, with an overlap of 50% in the three directions. Then, the patch-wise predicted output by the network is aggregated by stitching the patches together, to reconstruct the full-size segmentation result. For the airway segmentation experiments, we applied a lung mask to the final segmentation to remove any spurious noise prediction outside the lungs. For this operation, we use the same region growing lung mask as during training.

For the adversarial loss in Equation 3.2, the weight λ is set to 0.01 for all experiments in this work, based on visual inspection of the generated synthetic labels x_a .

3.3.5 Comparisons

We compared the results of our proposed method with the baseline 3D U-Net segmentation network described in Section 3.3.3 (U-Net baseline). Additionally, we compared our method with two previous refinement approaches: DoubleU-Net [96] and SCAN [97]. For both baselines, we reimplemented the methods from the original papers. The DoubleU-Net method consists of two consecutive U-Nets, with skip connections from the encoder of the first U-Net to the decoders of both U-Nets. The SCAN method uses a U-Net with a discriminator and adversarial loss, discriminating between the segmentation results and the ground truth. The weight for balancing the segmentation loss and the adversarial loss (low value on the adversarial term) is tuned between 0.001 and 0.1, on the validation sets for each application. For DoubleU-Net, no additional hyperparameters need to be tuned. Our implementations of DoubleU-Net and SCAN use the same 3D U-Net backbone as our proposed method and the first baseline.

We also conducted an ablation study of the proposed method (LR+Syn+LASN) by removing some of the components. We evaluated 1) a simple label refinement method with input data the original images and the initial segmentations without any synthetic errors (LR), 2) a label refinement method with synthetic errors added to the initial segmentations (LR+Syn(init)), and 3) a label refinement method with synthetic errors added to the ground truth segmentations but without the label appearance simulation network (LR+Syn).

3.3.6 Evaluation metrics

We evaluated the methods with the Dice coefficient to measure the overall segmentation quality, as well as with three metrics designed for tree-like structures: 1) centerline completeness, 2) centerline leakage, and 3) number of gaps. For the airway segmentation experiments, the required centerlines were obtained by applying a skeletonization method [89] to the ground truth segmentation mask. For the vessel segmentation experiments, the ground truth centerlines were manually annotated. The evaluation metrics are defined below:

Dice coefficient measures the voxelwise overlap between the predicted mask Y and the ground truth mask G :

$$Dice = \frac{2|Y \cap G|}{|Y| + |G|} \quad 3.4$$

Centerline completeness measures the proportion of the length of correctly detected centerlines (i.e., the intersection between the predicted mask Y and the ground truth centerlines G_{cl}) with respect to the length of ground truth centerlines G_{cl} :

$$Completeness = \frac{|Y \cap G_{cl}|}{|G_{cl}|} \quad 3.5$$

Centerline leakage measures the proportion of the length of false positive centerlines (i.e., the intersection between the predicted centerlines Y_{cl} and the ground truth background $1 - G$) with respect to the length of ground truth centerlines G_{cl} :

$$Leakage = \frac{|Y_{cl} \cap (1 - G)|}{|G_{cl}|} \quad 3.6$$

Gaps measures the number of continuity gaps in the correctly detected centerlines (i.e., the intersection between the predicted mask Y and the ground truth centerlines G_{cl}). It is calculated with connected component analysis [83] as follows:

$$Gaps = NCC(Y \cap G_{cl}) - NCC(G_{cl}) \quad 3.7$$

with NCC counting the number of 26-neighbor connected components in the input centerlines.

3.4 Results

3.4.1 Segmentation results

The results of our experiments for airway and brain vessel segmentation are shown in Tables 3.1 and 3.2, respectively. In both applications, the proposed label refinement method achieves the highest Dice and completeness scores, the lowest number of gaps, with a moderate leakage compared to the other methods. This indicates that our method succeeds in learning from the errors in the synthetic labels to correct errors in the real data. In both applications, the baselines with the highest completeness (SCAN for airways and DoubleU-Net for vessels) show a much higher leakage than our method. This indicates that these methods may lack the ability to learn relevant

Methods	Dice	Completeness	Leakage	Gaps
U-Net baseline [41]	0.76 (0.05)	0.74 (0.12)	0.23 (0.19)	95.73 (47.94)
DoubleU-Net [96]	0.77 (0.05)↑	0.73 (0.11)	0.21 (0.18)	99.93 (48.11)
SCAN [97]	0.77 (0.05)↑	0.75 (0.11)↑	0.31 (0.23)↓	98.83 (48.81)
LR	0.76 (0.05)	0.74 (0.11)↑	0.23 (0.17)	94.90 (47.66)
LR+Syn(init)	0.77 (0.06)	0.73 (0.12)	0.19 (0.17)	94.92 (50.14)
LR+Syn	0.79 (0.05)↑	0.73 (0.12)	0.17 (0.17)↑	93.54 (50.83)↑
LR+Syn+LASN (proposed)	0.79 (0.05)↑	0.75 (0.11)↑	0.20 (0.16)↑	91.63 (48.63)↑
LR+Syn+LASN +Unlabeled	0.81 (0.04)↑	0.77 (0.10)↑	0.19 (0.16)↑	90.53 (48.80)↑

Table 3.1: Results for airway segmentation. Average performance (standard deviation) over the results obtained from three random data splits. LR: simple label refinement network. LR+Syn(init): label refinement method with synthetic errors on initial segmentations. LR+Syn: label refinement method with synthetic errors on ground truth segmentations. LR+Syn+LASN: label refinement method with label appearance simulation network. ↑: significantly better than the U-Net baseline ($p < 0.05$). ↓: significantly worse than the U-Net baseline ($p < 0.05$). ↑↑: significantly better than the proposed method ($p < 0.05$). P -values are calculated by the paired two-sided Student’s T-test (on the average results from the three data splits). Boldface: best and not significantly different from the best results (semi-supervised results are not considered).

label structural information, and over-segment branches to increase the completeness rather than correcting errors in continuity.

In the ablation study, the label refinement method with synthetic errors (LR+Syn) achieves better Dice, leakage, and number of gaps scores than the baseline refinement network (LR), for both applications. For airway segmentation, the (LR+Syn) method has slightly lower completeness, while this is similar for vessel segmentation. Moreover, adding synthetic errors to the initial segmentations (LR+Syn(init)), in contrast to doing so to the ground truth segmentations (LR+Syn), achieves similar results in all metrics when compared to the baseline U-Net, for both applications. This suggests that the initial segmentations are too incomplete to add sufficient useful synthetic errors to train the refinement network. The proposed method, combining the synthetic errors and the label appearance simulation network (LR+Syn+LASN), achieves a much higher completeness, with similar Dice, leakage and number of gaps scores when compared to the method with only synthetic errors (LR+Syn), for both applications.

3.4.2 Semi-supervised results

We conducted experiments using semi-supervised learning to train the proposed label refinement method, to investigate the benefit of using additional unlabeled data for training. As labels in which to synthesize errors for the unlabeled data, we used segmentation results on the same data obtained by the proposed method (LR+Syn+LASN) trained on the labeled data. We denote these results as “pseudo

Methods	Dice	Completeness	Leakage	Gaps
U-Net baseline [95]	0.57 (0.10)	0.70 (0.18)	0.19 (0.18)	106.68 (161.41)
DoubleU-Net [96]	0.59 (0.09)↑	0.73 (0.18)↑	0.18 (0.16)	92.41 (151.27)↑
SCAN [97]	0.57 (0.09)	0.70 (0.18)	0.17 (0.15)	104.05 (160.91)
LR	0.57 (0.10)	0.70 (0.18)	0.16 (0.16)↑	82.05 (139.45)↑
LR+Syn(init)	0.58 (0.11)	0.71 (0.19)	0.18 (0.15)	69.91 (126.89)↑
LR+Syn	0.60 (0.11)↑	0.71 (0.19)	0.12 (0.11)↑	64.86 (115.21)↑
LR+Syn+LASN (proposed)	0.62 (0.10)↑	0.74 (0.20)↑	0.14 (0.11)↑	46.64 (76.57)↑
LR+Syn+LASN +Unlabeled	0.63 (0.09)↑	0.75 (0.18)↑	0.13 (0.11)↑	42.45 (71.26)↑

Table 3.2: Results for brain vessel segmentation. Average performance (standard deviation) over the results obtained from three random data splits. LR: simple label refinement network. LR+Syn(init): label refinement method with synthetic errors on initial segmentations. LR+Syn: label refinement method with synthetic errors on ground truth segmentations. LR+Syn+LASN: label refinement method with label appearance simulation network. ↑: significantly better than the U-Net baseline ($p < 0.05$). ↓: significantly worse than the U-Net baseline ($p < 0.05$). ↑↓: significantly better than the proposed method ($p < 0.05$). P -values are calculated by the paired two-sided Student’s T-test (on the average results from the three data splits). Boldface: best and not significantly different from the best results (semi-supervised results are not considered).

labels”. The error generation in these pseudo labels follows the same strategy and hyperparameters as in the previous experiments (Sections 3.2.3 and 3.3.2). The pseudo labels are also used as ground truth segmentations for the unlabeled images. These unlabeled data together with the labeled data in the previous experiments are then used to train a new label refinement network.

The results of our semi-supervised experiments for airway and brain vessel segmentation are shown in the last rows in Tables 3.1 and 3.2, respectively. Adding unlabeled data for training significantly improves the Dice score while the leakage remains similar, for both applications. For airway segmentation, the completeness is also improved, while this is similar for vessel segmentation. This suggests that for vessels, the labeled data provides enough information for the method to obtain segmentations with good completeness. For vessels, the number of gaps is also improved.

3.4.3 Influence of the synthetic error rate

The results of our experiments varying the maximum synthetic error rates (Section 3.3.2) are shown in Figure 3.4. For airway segmentation, with a smaller amount of “discontinuity” errors (0.1) the completeness is increased. Between 0.1 and 0.5, changing the amount of “discontinuity” errors in the synthetic labels does not affect much the method performance. In contrast, increasing the amount of “missing terminal branches” errors improves both Dice and completeness scores, reaching a peak when the maximum error rate is ≈ 0.75 . This supports our hypothesis that missing terminal

branches are relevant errors to be corrected in the initial airway segmentations. For vessel segmentation, a moderate amount (0.6) of “discontinuity” errors has a positive effect in the method performance.

When compared to the LR+Syn and LR+Syn(init) methods, the proposed label refinement method is able to learn from higher amounts of synthetic errors, thereby improving the label refinement performance.

3.5 Discussion

In this work, we proposed a novel label refinement method that can correct errors in the initial segmentations from a standard deep segmentation network such as the U-Net. The novelty of our method is that it uses labels augmented with realistic, synthetic errors as training samples, from where the label refinement network can learn to correct the errors. The synthetic errors are automatically generated to simulate common errors observed in the initial segmentations, and are then refined by a label appearance simulation network to resemble the appearance of real errors in the initial segmentations.

We evaluated our method on the segmentation of airways from chest CT scans and brain vessels from CTA images of the brain. In both applications, our method achieved significantly higher Dice overlap and completeness scores, with lower number of gaps and a comparable leakage, when compared to the baseline U-Net and other previous label refinement methods. When segmenting branching structures, a higher completeness means that more and/or longer branches are detected, especially the smaller ones which are challenging to segment.

The ability of our method to segment highly complete tree-like structures with more branches is clinically important, as this could lead to more sensitive biomarkers. For example, in airway analysis, the airway-artery ratio [17] and airway tapering [65] measures can be used to assess cystic fibrosis lung disease, and including more measurements from the smaller peripheral branches can allow earlier detection of the disease [110]. Moreover, the ability of our method to correct errors in continuity and thereby connect the segmentation is beneficial, as most methods to measure branches assume a fully connected segmentation and discard branches after a discontinuity.

The proposed method outperformed the state-of-the-art label refinement methods DoubleU-Net [96] and SCAN [97]. Moreover, using semi-supervised learning techniques to train our method with additional unlabeled data we can further improve the method performance, when compared to the fully supervised setting.

3.5.1 Comparison to other label refinement methods

The main difference between the proposed method and other label refinement methods is the use of a training dataset that includes labels augmented with synthetic errors. Instead of synthetic errors, the DoubleU-Net [96] method uses the original images masked by the initial segmentations to train the second network. Although the increased model capacity of DoubleU-Net may improve the segmentations, its ability to correct the errors may be limited by the fact that no new errors are introduced to the input of the second network. This makes it less efficient to implicitly exploit the label

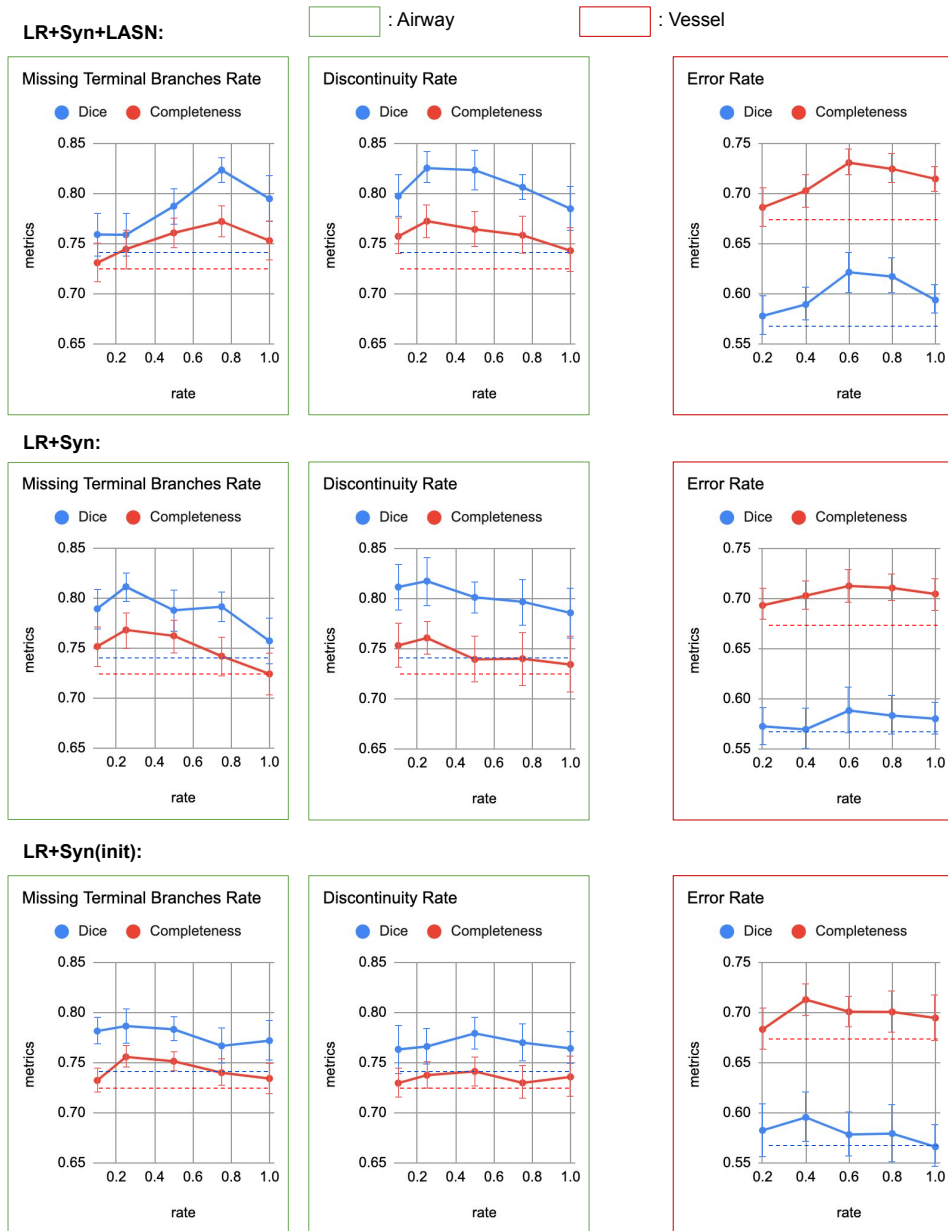


Figure 3.4: Influence of the hyperparameters of the proposed method, the maximum synthetic error rates, in the method performance, for airway and brain vessel segmentation. Results are shown as average performance with standard deviation (error bars), for Dice and completeness metrics, over three random data splits. The results for the baseline (LR) are displayed as dashed line.

structural information similar to a standard U-Net. The SCAN [97] method refines the segmentation by making it indistinguishable from the ground truth segmentation through an adversarial loss, where the distribution of the learned features may also provide the general structural information of the objects to be segmented. SCAN mainly focuses on simulating the appearance of the ground truth segmentations. However, SCAN is not designed to learn to correct structural errors explicitly, thus it may not capture the local continuity information as efficiently as our method. This is reflected by the significantly worse completeness reported by SCAN in Tables 3.1 and 3.2, for both applications. Our method provides an implicit way to enhance the network awareness of the structural information in the ground truth segmentations. For example, after seeing many errors in continuity, the refinement network is expected to understand the local continuity within elongated structures, and consequently to be able to correct these errors in the initial segmentations.

3.5.2 Synthetic errors for semi-supervised learning

With the proposed method, synthetic errors can be added to any pseudo labels obtained on unlabeled data, to be used in semi-supervised learning. In Section 3.4.2 we have shown that our method performance was significantly improved when using additional unlabeled data for training. Our approach to generate synthetic errors could be used together with other common semi-supervised approaches using pseudo labels, e.g., to optimize the prediction consistency of the same image from different models [111], or the prediction consistency of the same image with different transformations [112]. Using synthetic errors in these methods may improve the segmentation quality of pseudo labels from the unlabeled data, which could provide more informative features from these data and thereby improve the method performance.

3.5.3 Importance of realistic synthetic errors

The proposed label refinement network may underperform if the synthetic labels with errors used for training are too different from the initial segmentations. In our method, the synthetic errors are added to the ground truth segmentations, which have a fine and smooth appearance. In contrast, the initial segmentations are more irregular. Our proposed solution is to use a label appearance simulation network trained with an adversarial loss in order to make the appearance of the synthetic labels resemble that of the real initial segmentations. The results in Tables 3.1 and 3.2 clearly show the benefit of using the LASN network in our method. In both applications, without the LASN network could our method (LR+Syn) only slightly improve the segmentation performance with a reduced leakage, when compared to the baseline (LR). This may be due to the positive regularization effect of increasing the variety in the training data by including the synthetic labels with errors. Only after introducing the LASN network was our method able to improve the completeness while retaining an adequate leakage.

3.5.4 Applications to other segmentation tasks

The proposed label refinement method via error synthesis can be applied to other segmentation tasks. The core step is to identify common types of errors in the initial segmentations. For example, a common error we observed in prior work using the U-Net for the segmentation of the aorta and pulmonary arteries from chest CT scans [113] was that the segmentation of one of the structures often leaked into the other one, while being both independent anatomical structures. This is mostly due to the obscured boundaries of both arteries on the CT scan. This type of error can be simulated by locally removing the boundaries between the aorta and pulmonary artery classes. Applying our method to correct such errors may improve the overall segmentation performance for this application.

3.5.5 Limitations of the proposed method

The main limitation of the proposed method lies in the two-step design and implementation: 1) analyze the errors in the initial segmentations to identify the relevant types of errors, and 2) design and generate the synthetic errors based on these results. The first step requires observation and interpretation by the developers. The synthetic errors we used in this work are suitable for the segmentation of tree-like structures. However, the relevant types of errors generally differ across different applications and datasets, and therefore the synthetic errors we used are not directly applicable to other segmentation tasks. The second step is typically a complex image processing task. Nevertheless, once the synthetic errors are successfully designed for a given application, the training of our label refinement method can be done fully automatically.

A limitation of our validation of the proposed method is that we considered only two types of false negative errors (i.e., missing terminal branches and errors in continuity). We did not consider false positive errors because these were much less frequent in the initial segmentations and often appeared as disconnected blobs that can be easily removed without the need for more complex label refinement. Nevertheless, from the results obtained in this work we expect that our method can successfully correct other types of errors as well.

3.6 Conclusions

We presented a novel label refinement method that is able to learn from synthetic errors to refine the initial segmentations from a base segmentation network. A label appearance simulation network was applied to reduce the appearance difference between the synthetic labels and the real initial segmentations, thereby improving the generalizability of our method. On two segmentation tasks for branching structures, the proposed method achieved a significantly higher Dice overlap and centerline completeness, together with an improved continuity, when compared to previous label refinement methods. The segmentation performance of our method was further improved by using additional unlabeled data for training with semi-supervised learning techniques.

Acknowledgments

We want to thank Thirona and H. Tiddens, M. van de Corput and M. Bonte (Erasmus MC - Lung Analysis group) for providing the airway segmentations and anonymous chest CT data used in this work.

Author contributions

3

S. Chen: project lead, literature research, experiments, results analysis and writing. A. Garcia-Uceda: literature research, experiments, results analysis and writing. J. Su: literature research, experiments and results analysis. G. van Tulder: contribution to results analysis. L. Wolff: project data. T. van Walsum: supporting project manager, supervision of literature research, experiments, results analysis and writing. M. de Bruijne: lead project manager, supervision of literature research, experiments, results analysis and writing.

Availability of data and materials

The source code for our implementation of the proposed method is available here: <https://github.com/ShuaiChenBIGR/Label-refinement-network>

Ethical approval

This work involved medical images from human subjects acquired for clinical studies. Approval of the study protocols was granted by the central medical ethics committee of the hospital Erasmus MC, Rotterdam, the Netherlands (No. MEC-2013-338 and MEC-2014-235).

Funding

This work was partially funded by Chinese Scholarship Council (File No.201706170040), Netherlands Organization for Scientific Research (NWO) project VI.C.182.042, the MARBLE project (EFRO/OP-Oost: PROJ-00887), the Contrast project (Dutch Heart Foundation (CVON2015-01: CONTRAST), the Brain Foundation Netherlands (HA2015.01.06) and additional funding by the Ministry of Economic Affairs by means of the PPP Allowance made available by the Top Sector Life Sciences & Health to stimulate public-private partnerships (LSHM17016)).



CHAPTER 4

A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs



Abstract

This chapter presents an end-to-end deep learning segmentation method by combining a 3D U-Net architecture with a graph neural network (GNN) model. In this approach, the convolutional layers at the deepest level of the U-Net are replaced by a GNN-based module with a series of graph convolutions. The dense feature maps at this level are transformed into a graph input to the GNN module. The incorporation of graph convolutions in the U-Net provides nodes in the graph with information that is based on node connectivity, in addition to the local features learned through the downsampled paths. This information can help improve segmentation decisions. By stacking several graph convolution layers, the nodes can access higher-order neighborhood information without substantial increase in computational expense. We propose two types of node connectivity in the graph adjacency: 1) one pre-defined and based on a regular node neighborhood, and 2) one dynamically computed during training and using the nearest neighbor nodes in the feature space. We applied this method to the task of segmenting the airway tree from chest computed tomography (CT) scans. Experiments were performed on 32 CT scans from the Danish Lung Cancer Screening Trial dataset. We evaluated the performance of the U-Net-GNN models with two types of graph adjacency and compared it with the regular U-Net.

4.1 Introduction

Since recently, fully convolutional neural networks (CNN) are the state-of-the-art for many segmentation tasks [76], and in particular the U-Net architecture [56] for biomedical image segmentation. The U-Net consists of an encoding path, in which high-order features are extracted at several downsampled resolutions, followed by a decoding path, in which these features are leveraged to the full resolution to perform voxel-wise segmentation decisions. An extension of CNNs to graph structured data are graph neural networks (GNN) [114, 115], which have seen early application for segmenting structures that resemble graphs [116, 117]. Initial work of combining CNNs and GNNs was proposed by Shin et al. [117] for 2D vessel segmentation. This was a sequential pipeline in which the CNN was trained for feature extraction prior to applying the GNNs to learn global connectivity.

The segmentation of tree-like structures such as the airways from chest computed tomography (CT) scans is a complex task, with branches of varying sizes and different orientations while taking into account bifurcations. A comparison of airway extraction methods (prior to CNNs) was performed in the EXACT'09 challenge [38]. The results showed that all methods missed a significant amount of the small, peripheral branches. The U-Net has since been applied for airway segmentation in [45, 46, 77]. The 3D U-Net-based method in [77] is fully automatic and can segment the airways in a full lung in a single pass through the network. However, this method had problems to capture various small terminal branches. Moreover, GNNs have been applied to airway extraction as a graph refinement approach [116]. However, this method was not end-to-end optimized and relied on handcrafted features as input to the graph.

In this work, we present an end-to-end deep learning segmentation approach by combining the 3D U-Net with GNNs. This method replaces the two convolutional layers in the deepest level of the U-Net by a GNN-based module, comprising a series of graph convolutions. This end-to-end approach simultaneously learns local image features and global information based on graph neighborhood connectivity. We evaluated this U-Net-GNN method on the task of segmenting the airway tree from chest CT scans.

4.2 Methods

The proposed joint U-Net-GNN architecture is described in the following subsections. This approach integrates a GNN module at the deepest level of a 3D U-Net, and is schematically shown in Figure 4.1. The GNN module uses a graph structure obtained from the dense feature maps resulting from the contracting path of the U-Net. Each graph node can be viewed as a “supervoxel” from the downsampled regions with the corresponding vector of features. The connectivity of nodes in such a graph is described by the adjacency matrix and determines the neighborhood of each node when performing graph convolutions, as shown in Figure 4.2. The GNN module learns combinations of the input feature maps based on the graph topology, and outputs another graph with same “supervoxel” nodes as the input graph and the corresponding vector of learned features for each node. This GNN output are feature maps that are input to the upsampling path of the U-Net.

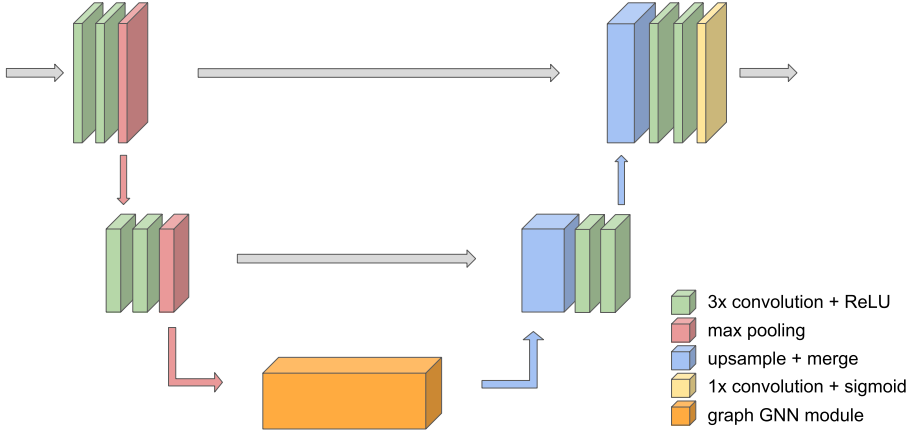


Figure 4.1: Schematics of the proposed U-Net-GNN model with 3 resolution levels.

We tested two types of node connectivity: 1) pre-defined and based on a regular neighborhood, with each node connected to its 26-direct neighbors, and 2) dynamically computed during training and based on choosing the nearest neighbors in the node feature space. Stacking several GNN layers allows nodes to access longer range information beyond the initial neighborhood, which can improve the segmentation decisions as more complex features that include relevant information from nodes far away in the 3D image can be used. This is in contrast with CNNs which rely on local feature extraction, and access long range information via successive convolutions and pooling, with a detriment of strong reduction in resolution. This long range access by the GNN module is useful when segmenting tree-like structures such as airways, as information from branches with similar features (e.g., shape, orientation) but that are further away in the 3D image can be used to segment a given branch. The computation of the dynamic graph adjacency is further explained in Section 4.2.2.

4.2.1 Graph neural network module

The main component of the GNN module is a series of graph convolutional layers [114, 115]. This operation can be seen as a generalization of the cartesian convolution to a graph setting. One of the formulations of graph convolution operation with separate processing of self connections is proposed in Kipf et al. [115], given by the equation,

$$\mathbf{H}^{(l+1)} = \sigma\left(\mathbf{H}^{(l)}\mathbf{W}_0^{(l)} + \mathbf{D}^{-1}\mathbf{A}\mathbf{H}^{(l)}\mathbf{W}_1^{(l)}\right) \quad 4.1$$

where $\sigma(\cdot)$ is the rectified linear unit activation function. $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times E}$ is the node feature matrix comprising N nodes and E features input to the l^{th} GNN layer. The learnable GNN filter weights are $\mathbf{W}_0^{(l)}, \mathbf{W}_1^{(l)} \in \mathbb{R}^{E \times E}$. $\mathbf{A} \in \{0, 1\}^{N \times N}$ is the binary

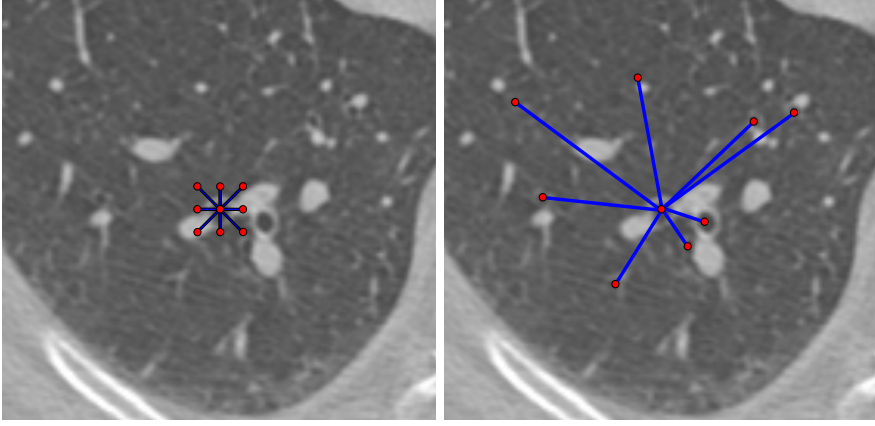


Figure 4.2: Illustration of the two types of node connectivity in the graph adjacency proposed for the GNN module, for a given voxel in the initial graph: regular and pre-defined (left) vs. irregular and dynamically computed (right).

adjacency matrix. \mathbf{D} is the degree matrix derived from \mathbf{A} with diagonal entries $D_{ii} = \sum_{j=1}^N A_{ij}, \forall i = 1 \dots N$. The adjacency matrix is largely sparse, with non-zero entries per node corresponding to the size of each node neighborhood (26 for the regular neighborhood case above). By processing the adjacency matrix as a sparse tensor, operations in Equation 4.1 are done efficiently.

The GNN module in the proposed method has L equal to 4 layers performing the operations in Equation 4.1 successively. By stacking several graph convolutions together, each node in the output graph updates its features with information from a higher-order neighborhood, which can improve the segmentation decisions. The initial graph features $\mathbf{H}^{(0)}$ are obtained as,

$$\mathbf{H}^{(0)} = f(\mathbf{Z}) \quad 4.2$$

where $\mathbf{Z} \in \mathbb{R}^{N \times F}$ is the F -dimensional node feature matrix derived from the U-Net and $f(\cdot)$ is a two-layered perceptron with F input units and E output units and rectified linear units, followed by a normalization layer. The transformation in Equation 4.2 allows for more complex representations of the input node features useful for the GNN module.

4.2.2 Irregular neighborhood

A GNN module with regular adjacency has limited node connectivity comprising only the direct neighbors. This constraint is imposed primarily to keep the adjacency matrix sparse to reduce the large memory footprint. In order to allow nodes in the graph to access information well beyond their directly connected neighbors, we propose an extension to the GNN module in Section 4.2.1 by using a graph adjacency with node connectivity as the k -nearest neighbors in the feature space. Node neighborhood is decided from the pairwise euclidean distance between nodes from \mathbf{Z} in the F -dimensional feature space. In this work, we set the number of neighbors k equal to 26.

The graph adjacency is dynamically computed during training for every input image in every epoch. This approach, we argue, enables the method to access irregular but meaningful neighborhoods and utilizes the inherent capabilities of using GNN-based learning over irregular neighborhoods in an improved fashion.

4.3 Experiments

4.3.1 Dataset

For our experiments, we used 32 low-dose chest CT scans from the Danish Lung Cancer Screening Trial [84]. All CT scans have a voxel resolution of roughly $0.78 \times 0.78 \times 1 \text{ mm}^3$. The reference segmentations are airway lumen extractions obtained from the method [58] applied on the union of the results of methods [32] and [33], and corrected by an expert observer.

4.3.2 Network implementation

The 3D U-Net upon which the U-Net-GNN approach is built derives from the method in [77] with 3 levels of resolution. Each level in the downsampling / upsampling path is composed of two $3 \times 3 \times 3$ convolutional layers with rectified linear (ReLU) activation, and followed by a $2 \times 2 \times 2$ pooling or upsampling layer, respectively. No padding is used in the convolutions in order to reduce the model memory footprint. There are 8 feature maps in the first level, and this is doubled or halved after every pooling or upsampling, respectively. The GNN module in the deepest level has twice as many output features per voxel as input units, i.e., $E = 2F$ in Equations 4.1 and 4.2. The final layer consists of a $1 \times 1 \times 1$ convolutional layer followed by a sigmoid activation function.

We apply a series of operations to extract input images from the chest CT scans. These are 1) cropping the CT scans to a bounding-box around pre-segmented lung fields, 2) extracting input image patches in a sliding-window fashion in axial dimension, and 3) applying image rigid transformation for data augmentation. The U-Net-GNN model used in our experiments is designed for the largest input images that can fit in GPU memory, for a GPU NVIDIA GeForce GTX 1080 Ti with 11 GB memory. This corresponds to an image size of $176 \times 352 \times 240$ in a batch containing only one image. The models were implemented using the Pytorch framework [82].

4.3.3 Training the models

The loss function used for training the models is the soft Dice loss [81],

$$\mathcal{L} = 1 - \frac{2 \sum_{x \in N_L} p(x)g(x)}{\sum_{x \in N_L} p(x) + \sum_{x \in N_L} g(x) + \epsilon} \quad 4.3$$

where $p(x), g(x)$ are the predicted voxel-wise airway probability maps and airway ground truth, respectively. The ground truth is masked to the region of interest (ROI), the lungs, indicated by the sub-index L , so that only voxels within this region contribute to the loss. This mask removes the trachea and part of the main bronchi

from the ground truth, so that training is focused on the smaller branches. The lung segmentation needed for this masking operation is easily computed by a region growing method in [32]. ϵ is a tolerance needed when there are no ground truth voxels in the image patch.

To train the networks, we used 16 CT scans for training, 4 CT scans for model validation and hyperparameter tuning, and the remaining 12 CT scans for testing. The Adam optimizer [88] is used with a learning rate that is chosen for each model as large as possible while ensuring convergence of the losses. This was 1×10^{-4} for U-Net-GNN models and 5×10^{-5} for U-Net models. All models were trained until convergence, and we retrieved the model with the overall minimum validation loss for testing. As convergence criterion, we monitored the moving average of the validation loss over 50 epochs, and training was stopped when its value 1) increases by more than 5%, or 2) does not decrease more than 0.1% over 20 epochs (patience). Training time was approximately 1–2 days on a GPU NVIDIA GeForce GTX 1080 Ti, while test time inference was less than 1 min per scan.

4.3.4 Experiments set-up

We evaluated four different models: two regular U-Nets with 3 levels (**UNetLev3**) and 5 levels (**UNetLev5**) of resolution, and two U-Net-GNN models each with 3 levels. The two U-Net-GNN models differ in the type of graph neighborhood used: 1) regular graph adjacency with 26 direct neighbors (**UGnnReg**), and 2) dynamic computation of the adjacency matrix (**UGnnDyn**), with 26 connections per node as described in Section 4.2.2. With our experiments we evaluated 1) the benefit of the GNN module at the deepest level of the U-Net when compared to the two convolutional layers of the regular U-Net, and 2) the difference in performance of the U-Net-GNN models when compared to a more complex U-Net like the UNetLev5.

The models are compared based on 1) Dice similarity coefficient, 2) airway completeness, 3) volume leakage, and two centerline distance error measures: 4) false negative d_{FN} and 5) false positive d_{FP} distances, as in [116]. Airway completeness is defined as the ratio of ground truth centerline length inside the predictions with respect to ground truth centerline full length. Volume leakage is defined as the ratio of number of false positive voxels with respect to the number of voxels in the ground truth. We refer the reader to [116] for the definition of the distance error measures d_{FN} and d_{FP} . The centerlines are obtained by applying skeletonization [89] to the prediction and ground truth masks. The trachea and main bronchi are removed in these measures from both the predictions and ground truth, similar to [38].

We computed the receiver operating characteristic (ROC) curves for all the models, with the mean airway completeness and volume leakage measured over the test set, and varying the threshold in the output probability maps used to obtain the final binary segmentations. Moreover, we computed the ROC operating point corresponding to a fixed mean volume leakage of 13% by estimating the correct threshold (with a minimum error of 1×10^{-4}). This threshold was, for the different models: UNetLev3 (0.1), UNetLev5 (0.04), UGnnReg (0.66), UGnnDyn (0.33). We used the resulting airway segmentations to evaluate the performance measures on the test set. To compare the models, we used the paired two-sided Student’s T-test on these performance measures.

We considered that a p -value lower than 0.01 indicates that the two sets of measures compared are significantly different.

The calculation of the dynamic graph adjacency as described in Section 4.2.2 is computationally expensive, both in runtime and memory. In order to fit the U-Net-GNN model in GPU memory and reduce its computational cost, we constrained the searching space of candidate nodes to compute the node connectivity to a cube area of maximum 5 voxels away from the given node. This limits the range for direct node connections, however long range information is still accessed via the four stacked graph convolutions as described in Section 4.2.1.

4.4 Results and discussion

4

The ROC curves for all the models are shown in Figure 4.3. The proposed U-Net-GNN models and the baseline UNetLev3 show similar results. In the detailed view, the UGnnReg shows a small improvement, i.e., higher airway completeness for a given volume leakage, over the baseline UNetLev3. Moreover, the UNetLev5 shows lower completeness than the other models for volume leakage higher than 11%, while the contrary occurs for lower leakage levels.

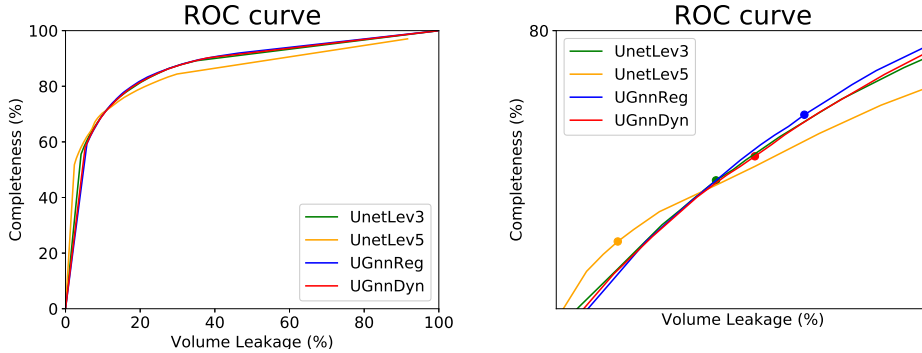


Figure 4.3: ROC curves for all the models, computed with the mean airway completeness and volume leakage measured over the test set, and varying the threshold in the output probability maps. Right: detailed view, showing the ROC operating points for all models corresponding to a threshold of 0.5.

Performance measures on the test set for all the models are shown in Figure 4.4, corresponding to an ROC operating point with mean volume leakage 13%. In Dice coefficient, airway completeness and volume leakage, there is no significant difference between all the models ($p > 0.1$). Nevertheless, in centerline distance error measures, the proposed U-Net-GNN models show significantly lower d_{FN} and higher d_{FP} with respect to the baseline UNetLev3: in d_{FN} , comparing UGnnDyn ($p = 0.001$) and UGnnReg ($p < 0.01$) with UNetLev3; and in d_{FP} , comparing both U-Net-GNN models with UNetLev3 ($p < 0.001$). Moreover, the UNetLev5 shows no significant difference in d_{FP} and d_{FN} with the other models ($p > 0.1$).

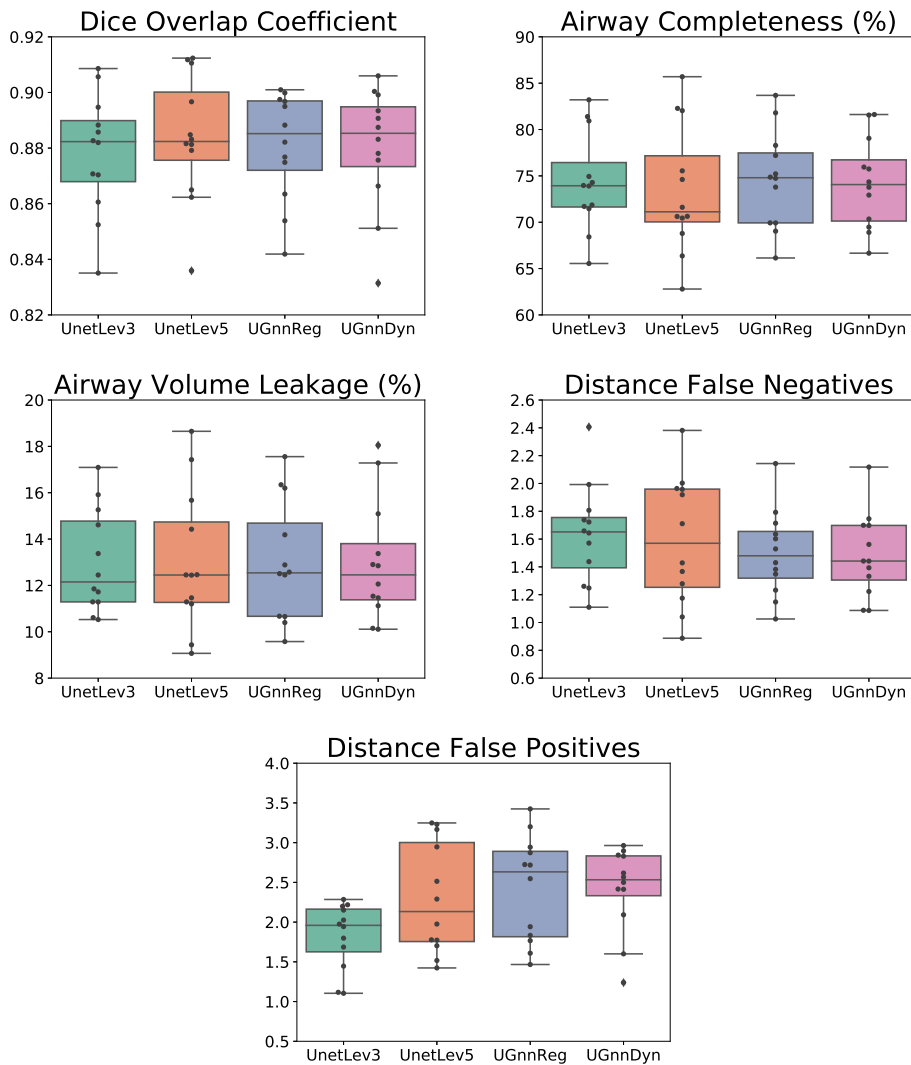


Figure 4.4: Performance measures on the test set for all the models, corresponding to an ROC operating point with mean volume leakage 13%. From left to right: 1) Dice similarity coefficient, 2) airway completeness, 3) volume leakage, 4) distance false negative error d_{FN} , and 5) distance false positive error d_{FP} . For each boxplot, the box shows the quartiles of the data (defined by the median, 25% and 75% percentiles), the whiskers extend to include the data within 1.5 times the interquartile range from the box limits, and the markers show all the data points.

The small and significant improvement in distance false negative error d_{FN} , with an on par Dice score and airway completeness, indicates that the proposed U-Net-GNN models can segment slightly more complete airway trees, when compared to the two U-Net models, with more and/or longer peripheral branches. This is because the centerline distance error measure d_{FN} is not dependent on the scale of airways and provides a uniform measure of accuracy in detecting branches. In contrast, the airway completeness measure is a binary evaluation of branches with respect to the reference, and small improvements in centerline detection may not be reflected in the overall score. The more complex model UNetLev5 does not show any difference in any measure over the baseline UNetLev3. Moreover, the U-Net-GNN models have fewer trainable parameters ($\approx 50k$) than the UNetLev3 ($\approx 90k$) and much less than the UNetLev5 ($\approx 1.4M$).

One aspect that might limit the performance of the proposed U-Net-GNN models is that the GNN module is operating only on the deepest level of the U-Net, where the input images have undergone three downsampling operations. A more powerful U-Net-GNN model can be formulated by replacing all skip connections in the U-Net with separate GNN modules, each operating at that image resolution level. However, we did not experiment with this model design due to the large memory footprint required, which exceeded the available GPU resources at our disposal.

4.5 Conclusions

We presented a joint U-Net-GNN-based segmentation method with an application to segment airways from chest CT scans. By introducing a GNN module with graph convolutions at the deepest level of the U-Net, the proposed method is able to learn and combine information from a larger region of the image. The proposed U-Net-GNN models show a small and significant improvement in the centerline distance false negative error measure over the regular U-Net, with on par Dice score and airway completeness, for a fixed mean volume leakage. This indicates that the proposed U-Net-GNN models can segment slightly more complete airway trees. Moreover, this is achieved with fewer trainable parameters.

Author contributions

A. Garcia-Uceda: project lead, literature research, experiments, results analysis and writing. R. Selvan: contribution to literature research, results analysis and writing. Z. Saghir: project data. M. de Bruijne: lead project manager, supervision of literature research, experiments, results analysis and writing.



CHAPTER 5

Creating a training set for artificial
intelligence from initial segmentations
of airways



Abstract

Airway segmentation is important for research about pulmonary disease, but require a large amount of time by trained specialists. We used an openly available software to improve airway segmentations obtained from an artificial intelligence (AI) tool, and retrained the tool to achieve a better performance. Fifteen initial airway segmentations from low-dose chest computed tomography scans were obtained with a 3D U-Net AI tool previously trained on Danish Lung Cancer Screening Trial and Erasmus-MC Sophia datasets. Segmentations were manually corrected in 3D Slicer. The corrected airway segmentations were used to retrain the 3D U-Net. Airway measurements were automatically obtained, and included airway count, airway length and luminal diameter per generation from the segmentations. Correcting segmentations required 2–4 hours per scan. Manually corrected segmentations had more branches ($p<0.001$), longer airways ($p<0.001$) and smaller luminal diameters ($p=0.004$) than initial segmentations. Segmentations from the retrained 3D U-Net trended towards more branches and longer airways compared to the initial segmentations. The largest changes were seen in airways from 6th generation onwards. Manual correction results in significantly improved segmentations and is potentially a useful and time-efficient method to enhance the AI tool performance on a specific hospital or research dataset.

5.1 Introduction

Airway segmentation from chest computed tomography (CT) scans is important in the study of pulmonary disease such as chronic obstructive pulmonary disease (COPD) [118]. High-quality airway segmentation datasets are difficult to create, yet they are necessary for the training of artificial intelligence (AI) tools. Manually segmenting airways from noisy low-dose CT scans is time consuming and error prone. Automatic methods that can provide an adequate segmentation of large airways via region growing may fail and require manual correction [38, 119].

The volume of thoracic CT scans in clinical care will increase due to a growing burden of respiratory diseases and the introduction of imaging-based cancer screening [120]. Computer assistance will become increasingly important in the radiology workflow. This should be supplemented with robust AI tools that can increase the accuracy and speed of the diagnosis. Medical datasets used to train AI tools are typically small, due to the limited availability of imaging data and ground truth annotations. In contrast, there is a wide range in possible CT scanning and population characteristics. Thus, pre-trained AI tools have issues generalizing when tested on new data, with typically different characteristics. In this scenario, the need for quickly adapting existing AI models trained on different data may prove very useful.

AI segmentation tools are being widely studied for their potential in automation, accuracy and reliability. However, their usability comes at the cost of flexibility inherent in AI systems. To achieve the highest accuracy, AI tools require training on CT scans that are similar to those it will be applied on. X-ray tube current, voltage, reconstruction methods and other parameters change the resulting CT image and may have an impact on segmentation performance [3].

So far, the methodology for obtaining high-quality ground truth segmentations of airways using openly available software is lacking. While many airway segmentation tools already exist, those that provide a highly detailed segmentation may be only available for sale, run as a service or are tied to specific CT scanner brands and hospital / research set-ups [44, 75].

In this work, we propose a solution to generate good ground truth airway segmentations by improving initial segmentations obtained using openly available AI software. We also assessed the change in AI performance on our low-dose chest CT protocol after retraining the AI tools using the corrected segmentations [121].

5.2 Methods

5.2.1 Initial segmentations

We used a 3D U-Net method designed for automatic airway segmentation [41]. The 3D U-Net is a deep learning model for biomedical image segmentation, which classifies image voxels as airway / non-airway. The image filters in the convolution layers of the method were optimized automatically using training images and reference segmentations. For all our experiments, we used the same U-Net layout and hyperparameters as in [41], which are well-suited for airway segmentation.

The 3D U-Net was trained on Danish Lung Cancer Screening Trial (DLCST) [84] and Erasmus MC-Sophia data (ErasmusMC) (pediatric patients with cystic fibrosis) [17]. This model was used to obtain the initial airway segmentations from CT scans of fifteen randomly selected participants from the Imaging in Lifelines (ImaLife) study [121]. The CT scans used were low-dose unenhanced, obtained using a 16-slice CT scanner (SOMATOM Sensation 16, Siemens Medical Solutions) with a pitch of 3 (with FOV 350) or 2.5 (with FOV 400) and 1 mm increments at a tube voltage of 120 kVp and reference current of 20 mAs [122]. Images were reconstructed with overlapping 0.7 mm increments using the Qr59 kernel. The ImaLife study is part of the northern Netherlands study and includes participants of at least 45 years of age from the general population. Complete details on ImaLife patient characteristics can be found in Table 5.1 and in the referenced material [121]. Differences in population and scanning parameters for DLCST and ErasmusMC datasets compared to ImaLife dataset contributed to incomplete initial airway segmentations. The prediction threshold of the 3D U-Net probability maps was set to 0.5, which resulted in a low number of false positive airways in the initial segmentations. Thus, most manual corrections required addition of missing branches, rather than removal of false branches.

Participants	Number or Mean \pm SD
Males	7
Females	8
Age (years)	57.3 \pm 6.3
Height (cm)	174.1 \pm 7.6
Weight (kg)	74.67 \pm 12.1
Current smokers	7

Table 5.1: Population characteristics of the ImaLife CT scans included in the study (n=15). SD: standard deviation.

5.2.2 Manual correction of segmentations

The initial airway segmentations were imported into 3D Slicer 4.1 (<https://www.slicer.org>) [123]. Window settings were set to a width of 800 and a level of -625 to better visualize the airway lumen. One medical doctor with 6 months of work and training in pulmonology (I.D.) performed the manual corrections of the segmentations.

The workflow screen displayed the coronal, sagittal, transverse and three-dimensional (3D) views, as shown in Figure 5.1. Corrections were performed using the segment editor tool in 3D Slicer [123]. The binary segmentation provided by the 3D U-Net was imported into the segment editor. Next, the airways segmentations were completed using the paint tool, with a spherical brush and brush size dynamically set to 1–3% of the active window size, based on the size of the airway. 3D Slicer provides tools to follow along an incomplete airway in the 3D view and identify it on the three views. In this manner, it was possible to quickly complete airway segmentations as they were identified on all three orientations simultaneously, with the results instantly visible on the 3D view.

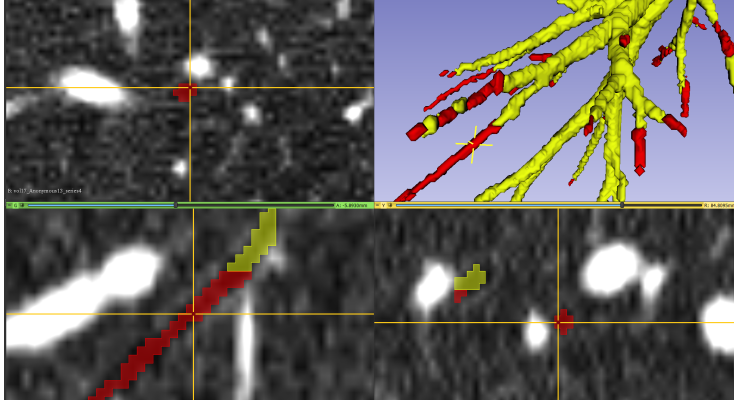


Figure 5.1: A 3D Slicer workspace for fast identification and correction of incomplete airways. Yellow: Incomplete airway segmentation of an ImaLife participant. Red: manual correction of the airways.

The initial airway segmentations were combined with the manual corrections and exported as a set of DICOM slices.

5.2.3 3D U-Net evaluation

We used 15 ImaLife CT scans and the corrected airway segmentations to train a new 3D U-Net, referred to as “retrained” model. For training and evaluation, we used a 5-fold cross-validation setting, splitting the dataset into 5 groups of equal size, and training 5 different models. For each model, one split group is assigned as testing set, and the remaining 4 groups are used as training set. Within each training fold, 83% of data is used for model weight optimization, and the remaining 17% for model selection. We evaluated each trained model on their corresponding independent testing set. Each training fold contains 12 CT scans. Despite the small number of scans, the used 3D U-Net [41] was validated with varying sizes of training sets and the learning curves showed good performance with similar numbers of CT scans.

To assess the change in AI performance when introducing a larger heterogeneous dataset, we trained a second model with a combination of ImaLife, DLCST and ErasmusMC data, referred to as “combined” model. We used the same 5-fold cross-validation split of the ImaLife data as for the “retrained” model above, and added 20 CT scans each from DLCST and ErasmusMC data to the training folds. Trained models were used to segment airways from the ImaLife CT scans for comparison with the initial segmentations. The overall process is summarized in the flowchart shown in Figure 5.A.1 in the Appendix.

5.2.4 Analysis of segmentations and statistical analysis

From the airway segmentations obtained by the 3D U-Net, branches and their generation number were extracted automatically, similarly to the evaluation used in the EXACT’09 challenge [38]. The airway generation is defined as the number of branch

bifurcations counted in the path linking the given branch and the first branch in the airway tree, i.e., the trachea. Thus, the trachea is generation 0, main bronchi generation 1, etc. Automatic measurements of lumen diameter were extracted every 1 mm along the centerline of and averaged per branch. The branch length was calculated as the distance between bifurcations along the branch centerline.

Comparison were made between the initial segmentations and segmentations from the retrained and combined models trained with the manually corrected segmentations. Results were analyzed using Python (Python Software Foundation, <https://www.python.org/>) and the SciPy package [124]. We used the Wilcoxon signed-rank test with Bonferroni correction for analysis. All comparisons were with respect to the initial, incomplete segmentations. A p -value lower than 0.05 was considered significant.

5.3 Results

5.3.1 Segmentations

Fifteen ImaLife CT scans were segmented by the 3D U-Net. The obtained initial airway segmentations were manually corrected, as shown in Figure 5.2. In two cases of large mucus plugging, the 3D U-Net continued to segment the airways beyond the blockage without the need for manual interaction, as shown in Figure 5.3. The time to complete a manual correction ranged from 2 to 4 hours.

5.3.2 Airway count

The initial, incomplete segmentations had the lowest count of airways 151 (interquartile range (IQR) 131–169). This was followed by the retrained model segmentation with 170 airways (IQR 161–197) ($p=0.098$, initial vs. retrained), and the combined model segmentation with 174 airways (IQR 146–201) ($p=0.089$, initial vs. combined). The manually corrected segmentation had the highest count of airways with 179 (IQR 167–215) ($p<0.001$, initial vs. manual). These results are shown in Figure 5.4a, and in Table 5.2. The largest differences were seen in airways from 6th generation onwards, as shown in Figure 5.A.2 in the Appendix.

5.3.3 Airway length

Airway length increased with manual correction and retraining. The initial segmentation had a total airway length of 2319 mm (IQR 1905–2588 mm) which was the lowest among all segmentations. This was followed by the combined model segmentation with total length 2561 mm (IQR 2309–3067 mm) ($p=0.079$, initial vs. combined), and the retrained model segmentation with total length 2622 mm (IQR 2296–3345 mm) ($p=0.051$, initial vs. retrained). The manually corrected segmentation had the highest total length of 2917 mm (IQR 2508–3492 mm) ($p<0.001$, initial vs. manual). These results are shown in Figure 5.4b, and in Table 5.2. Airways from the 6th generation onwards showed the largest differences, as shown in Figure 5.A.3 in the Appendix.

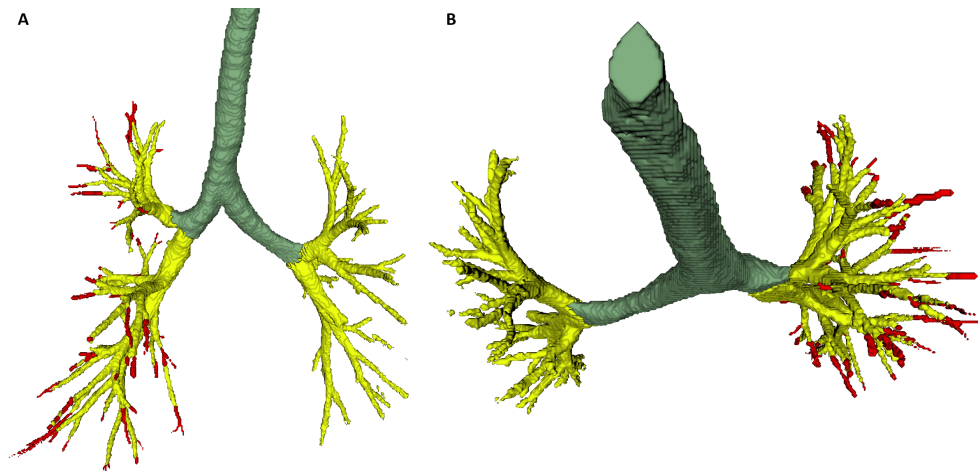


Figure 5.2: Example of an incomplete segmentation of the airway tree for an ImaLife participant (in yellow), and a manually corrected segmentation (in red) of the right lung.

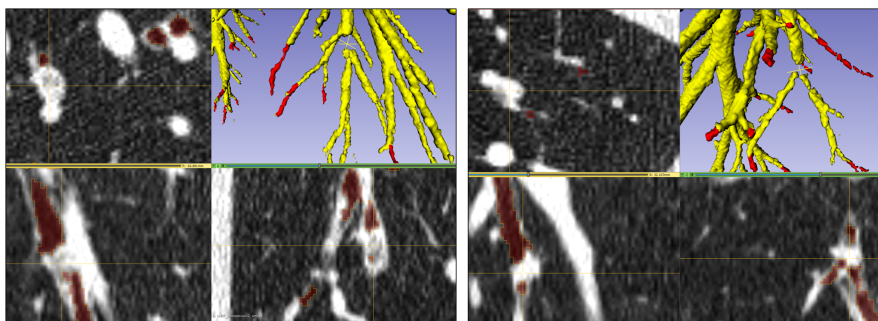


Figure 5.3: Two examples of large mucus plugging with total focal occlusion of the airway of an ImaLife participant. The 3D U-Net completed segmentation of branches distal to the occlusion without supervision.

5.3.4 Airway lumen diameter

Relative to the initial segmentation with mean airway luminal diameter of 5.5 mm (IQR 5.0–5.9 mm), the lumen diameters decreased with manual correction to 5.3 mm (IQR 4.9–5.6 mm) ($p=0.009$, initial vs. manual), and with the retrained model segmentation to 4.9 mm (IQR 4.7–5.5 mm) ($p=0.004$, initial vs. retrained). There was no significant difference between the mean luminal diameter of the initial segmentation and the combined model segmentation with 5.0 mm (IQR 4.6–6.1 mm) ($p=0.172$, initial vs. combined). These results are shown in Figure 5.4c, and in Table 5.2. A detailed breakdown per generation is shown in Figure 5.A.4 in the Appendix.

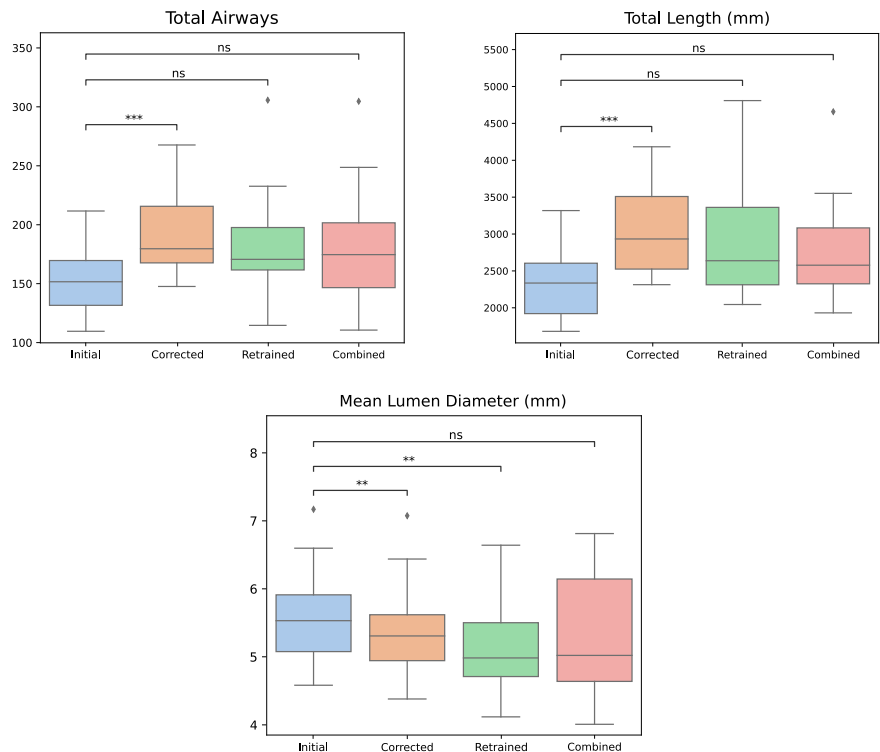


Figure 5.4: Boxplots showing 1) total airway count, 2) total airway length and 3) mean luminal diameter, for the segmentation results obtained with the retrained and the combined 3D U-Nets. ns: not significant, $*p<0.05$, $**p<0.01$, $***p<0.001$.

Measurements	Count	Length (mm)	Diameter (mm)
Initial	151 (131 – 169)	2319 (1905 – 2588)	5.5 (5.0 – 5.9)
Corrected	179 (167 – 215)	2917 (2508 – 3492)	5.3 (4.9 – 5.6)
Retrained	170 (161 – 197)	2622 (2296 – 3345)	4.9 (4.7 – 5.5)
Combined	174 (146 – 201)	2561 (2309 – 3067)	5.0 (4.6 – 6.1)

Table 5.2: Summary of measurements for each set of segmentations (n=15). Data is displayed as median (interquartile range).

5.4 Discussion

We outlined the process for correcting airway segmentations from initial, incomplete segmentations on low-dose CT scans for the purpose of training AI tools. Manual correction resulted in significantly more complete airway segmentations, and retraining the 3D U-Net resulted in improved segmentations, with the greatest changes seen from the 6th generation onwards. Notably, small airways play an important role in lung diseases such as asthma, COPD, and cystic fibrosis, and their accurate

detection is important for an accurate diagnosis and sensitive monitoring of respiratory diseases [110, 125]. A focus on improving the segmentation of smaller airways could therefore help in the research of bronchial parameters of early disease [126]. With the proposed method, it is possible to quickly improve airway segmentations and retrain an AI model.

The research for robust bronchial parameters sometimes includes the evaluation of aggregate measures, such as the total airway count and airway tapering [127, 128]. If these measures are obtained from incomplete segmentations, the summary measure may be incorrect. This is illustrated in our study by the decrease in median lumen diameter after correction and retraining. The initial segmentations included too much of the lumen wall and did not include enough of the smaller airways that were visible on the CT scan. This resulted in a significantly larger median airway lumen aggregate measure.

A major challenge for training AI models in radiology is that often only small, specific datasets from a narrow range of scanning parameters and population characteristics are available for training AI models, and the ground truth segmentations are difficult and tedious to obtain (for airways, this takes up to 15 hours to complete for one patient [17]). This makes the design of AI tools that generalize well to data from a broader range of scanning parameters and population characteristics very difficult. In turn, pre-trained AI segmentation models may fail when applied on data dissimilar to their training data. Several AI airway segmentation tools have been reported in the literature, which are typically trained and tested on their own in-house datasets and reference segmentations [32, 34]. However, when deploying the trained AI models on other data with different characteristics and scanning parameters, their performance may drop drastically [129]. Retraining with use-case specific data allows for the use of AI models in institutions with different scanning techniques.

The aim of adding the DLCST and Erasmus MC-Sophia datasets for training was to improve the AI performance with heterogeneous data, as the DLCST scanning protocol differs slightly to that from ImaLife, and Erasmus MC-Sophia includes pediatric patients with cystic fibrosis. However, the combined AI model did not significantly improve the AI performance for ImaLife CT scans.

A strength of the proposed methodology is the use of openly available software. While this work focuses on airway segmentation, the same methodology can be used to optimize potentially any other AI segmentation model. Moreover, it is much less time costly than generating fully manual airway reference segmentations.

A limitation of our study is that we investigated only one dataset, with a small sample size and based on low-dose CT scans acquired at high-pitch in a general adult population. Nevertheless, the used 3D U-Net can operate on small datasets, as shown by the learning curves where a model trained with only 14 images had just slightly lower performance than a model trained with 28 images [41]. Another limitation is that the manual corrections were performed by only one researcher, and we did not assess the impact of inter-observer variability on the completeness of the corrected segmentations.

5.5 Conclusions

We showed that openly available software can be used to manually correct initial, incomplete airway segmentations with significant improvement. The resulting segmentations can be used to retrain AI models to increase their efficacy for different scanning protocols and applications. This allows for the quick creation of datasets for training AI models that match their use case.

Author contributions

I. Dudurych: project lead, experiments, airway corrections, statistical analysis and writing. A. Garcia-Uceda: U-Net design and training, and writing. Z. Saghir: project data, methodology. H. Tiddens: project data, methodology. R. Vliegthart: supporting project manager, supervision of methodology, project data and writing. M. de Bruijne: lead project manager, supervision of methodology, project data and writing.

5

Availability of data and materials

The 3D U-Net used for this research is openly available on github and includes a model file trained on DLCST and Erasmus-MC Sophia datasets: <https://github.com/antoniooguj/bronchinet>

Ethical approval

Institutional Review Board approval was obtained. The ImaLife study was registered with the Dutch Central Committee on Research Involving Human Subjects (<https://www.toetsingonline.nl>, Identifier: NL58592.042.16). Written informed consent was obtained from all subjects (patients) in this study.

Funding

This work is part of the first author's PhD which has received funding by IMDI (Innovative Medical Devices Initiative) / ZonMw (Netherlands Organization for Health Research and Development).

Appendix

5.A Additional figures and tables

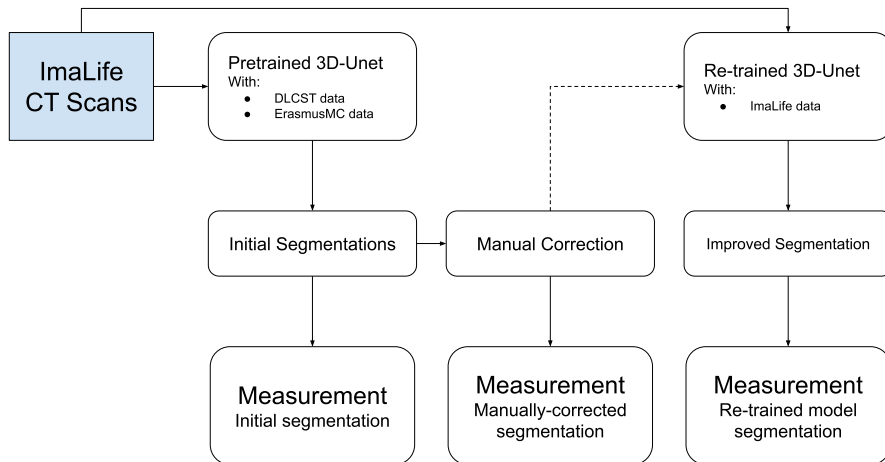


Figure 5.A.1: Flowchart demonstrating the segmentation, retraining and measurement process.

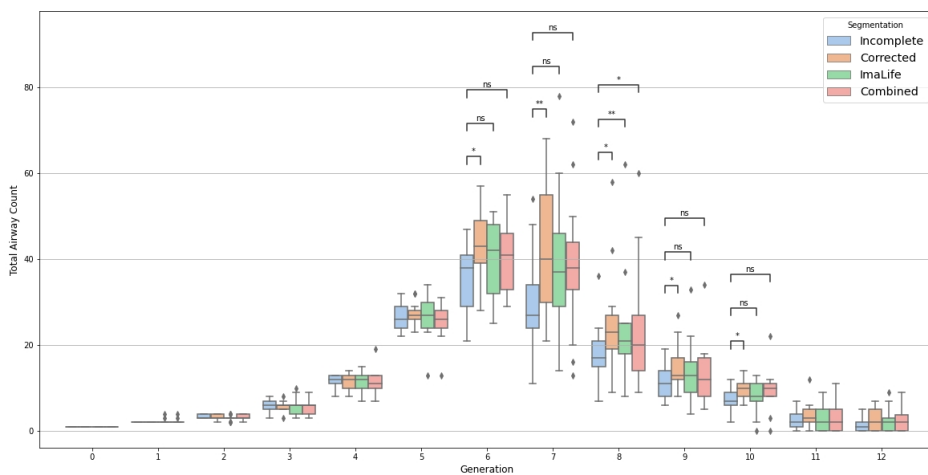


Figure 5.A.2: Boxplot of total airway count per generation. ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

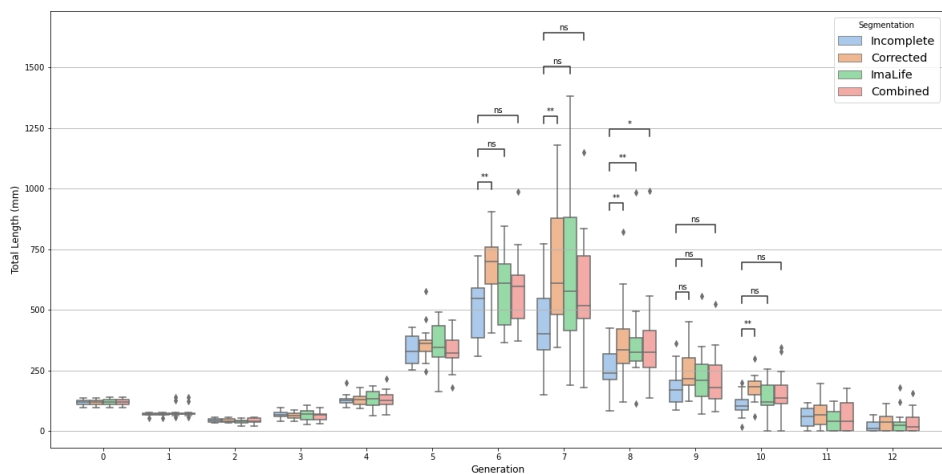


Figure 5.A.3: Boxplot of total airway length per generation. ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

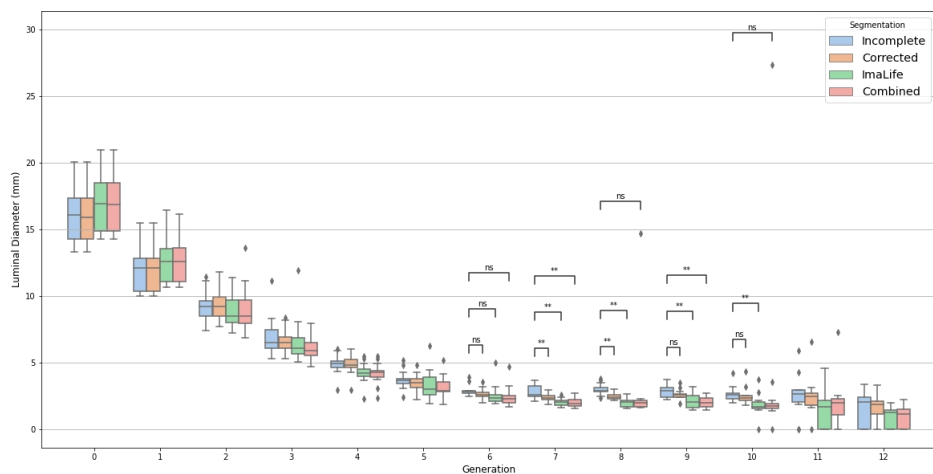


Figure 5.A.4: Boxplot of mean luminal diameter per generation. ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



CHAPTER 6

Reproducibility of a combined
artificial intelligence and
optimal-surface graph-cut method to
automate bronchial parameter
extraction



Abstract

Computed tomography (CT)-based bronchial parameters correlate with disease status. Segmentation and measurement of the bronchial lumen and walls usually requires significant manpower. We evaluated a combined method of deep learning and optimal-surface graph-cut to automatically segment the airway lumen and wall from chest CT scans and calculate bronchial parameters. A deep learning model designed for automatic airway extraction was trained on 24 Imaging in Lifelines (ImaLife) low-dose CT scans. This model was combined with an optimal-surface graph-cut method for airway wall segmentation. These tools were used to calculate bronchial parameters from CT scans of 188 ImaLife participants who had two scans an average of 3 months apart. Bronchial parameters were compared for reproducibility assessment, assuming no change between scans. 374/376 CT scans were successfully measured. Segmented airway trees contained a mean of 10 generations and 250 branches. Coefficient of determination (R^2) ranged from 0.93 at the trachea to 0.68 at the 6th generation, dropping to 0.51 at the 8th generation, for luminal area (LA). Corresponding values for wall area percentage (WAP) were 0.86, 0.67 and 0.42, respectively. Bland-Altman analysis of LA and WAP demonstrated mean difference close to 0 per generation and limits of agreement (LoA) ranging ± 10 –23% of the overall measurement range per generation for each parameter. From 7th generation on, there was a sharp decrease in reproducibility of measurements and a widening of LoA spread. The proposed approach for automatic bronchial parameter measurement on low-dose chest CT scans is a reliable way to assess the airway tree down to the 6th generation.

Based on: I. Dudurych*, A. Garcia-Uceda*, J. Petersen, Y. Du, R. Vliegenthart, and M. de Bruijne, “Reproducibility of a combined artificial intelligence and optimal-surface graph-cut method to automate bronchial parameter extraction,” *Submitted*

* indicates equal contributions



CHAPTER 7

The effect of CFTR modulators on
structural lung disease in cystic
fibrosis



Abstract

Newly developed quantitative chest computed tomography (CT) outcomes designed specifically to assess structural abnormalities related to cystic fibrosis (CF) lung disease are now available. CFTR modulators potentially can reduce some structural lung abnormalities. We aimed to investigate the effect of CFTR modulators on structural lung disease progression using different quantitative CT analysis methods specific for people with CF (PwCF). PwCF with a gating mutation (ivacaftor) or two Phe508del alleles (lumacaftor-ivacaftor) provided clinical data and underwent chest CT scans. Chest CT scans were performed before and after initiation of CFTR modulator treatment. Structural lung abnormalities on the CT scan were assessed using the Perth Rotterdam Annotated Grid Morphometric Analysis for CF (PRAGMA-CF), airway-artery (AA) dimensions, and CF-CT methods. Lung disease progression (0 to 3 years) in exposed and matched unexposed subjects was compared using analysis of covariance. To investigate the effect of treatment in early lung disease, subgroup analyses were performed on data of children and adolescents aged <18 years. We included 16 modulator exposed PwCF and 25 unexposed PwCF. Median (range) age at the baseline visit was 12.55 (4.25 to 36.49) years and 8.34 (3.47 to 38.29) years, respectively. The change in PRAGMA-CF %airway disease (-2.88 (-4.46,-1.30), $p=0.001$) and %bronchiectasis extent (-2.07 (-3.13,-1.02), $p<0.001$) improved in exposed PwCF compared to unexposed. Subgroup analysis of pediatric data showed that only PRAGMA-CF %bronchiectasis (-0.88 (-1.70,-0.07), $p=0.035$) improved in exposed PwCF compared to unexposed. In this preliminary real-life retrospective study CFTR modulators improve several quantitative CT outcomes. A follow-up study with a large cohort and standardization of CT scanning is needed to confirm our findings.

Based on: L. Mok*, **A. Garcia-Uceda***, M. Cooper, M. Kemner van de Corput, M. de Bruijne, N. Feyaerts, T. Rosenow, K. De Boeck, and S. Stick, "The effect of CFTR modulators on structural lung disease in cystic fibrosis," *Submitted*

* indicates equal contributions



CHAPTER 8

Quantitative chest computed
tomography scoring technique for
bronchiectasis (BEST-CT)



Abstract

This chapter presents a newly developed computed tomography (CT) scoring system (BEST-CT) to quantify and phenotype structural lung disease in bronchiectasis (BE) patients. Recent CT scans of BE patients with chronic *Pseudomonas aeruginosa* infection enrolled in the iBEST study were collected and analyzed with the BE scoring technique for CT (BEST-CT). BEST-CT scores are expressed as % of total lung volume. Scoring items are: consolidation / atelectasis, BE with mucus plugging (BEMP), BE without mucus plugging (BEwMP), airway wall thickening (AWT), mucus plugging (MP), ground-glass opacities, emphysema / bullae, healthy airways (HA), and healthy parenchyma (HP). Furthermore, composite scores total bronchiectasis ($\%TBE = \%BEMP + \%BEwMP$), airway disease ($\%AD = \%TBE + \%MP + \%AWT$) and total disease ($\%DIS =$ all but $\%HA$ and $\%HP$) were calculated. Intra-class correlation coefficients (ICC) were calculated. BEST-CT scores were compared with the Hartmann CT scoring method, intra-branch tapering, FEV_1 , bronchiectasis severity index (BSI), quality of life (QOL), and exacerbations. 84 inspiratory CT scans were included. BEST-CT subscores were (median (interquartile range)): $\%TBE$ 3.0 (1.4–5.1), $\%AD$ 6.5 (3.4–11.8), and $\%DIS$ 9.4 (6.0–17.7). ICC values for $\%TBE$, $\%AD$ and $\%DIS$ were good to excellent. Significant correlations were found between $\%TBE$ and Hartmann BE scores ($r=0.69$, $p<0.001$), FEV_1 ($r=-0.24$, $p=0.027$), and intra-branch tapering ($r=-0.34$, $p=0.002$). No significant correlations were found between BEST-CT scores and BSI, QOL, or exacerbations. The BEST-CT is a reproducible CT scoring system to phenotype BE patients, that correlates with other CT parameters and spirometry. Using this method, considerable heterogeneity of lung disease was observed.



CHAPTER 9

Discussion



9.1 General discussion

A general discussion of the main findings and achievements of the research presented in this thesis is given below:

1. Deep learning methods, and particularly convolutional neural networks, are a powerful and robust methodology to segment the bronchial tree from chest CT scans.

In Chapter 2 we showed that our method based on the deep learning U-Net architecture obtains high-quality airway segmentations in various datasets. This included challenging cases such as 1) CT scans with a wide range of airway abnormalities, and 2) CT scans that are dissimilar to those used for training the method. Moreover, we used relatively small training datasets composed of 20–30 CT scans and yet our method obtained good performance. We believe this is partly because our method uses as backbone the original U-Net with minimal modifications. This makes our method simple and with a reduced number of trainable parameters, which helps prevent overfitting to the training data when using small training sets. Thanks to this, our method is more likely to generalize well to CT data that are dissimilar to the training data. This was illustrated in the evaluation on the public EXACT’09 dataset in Chapter 2, where our method was among the best performing ones when compared to other deep learning segmentation methods, while using smaller training datasets composed of CT data that are dissimilar to EXACT’09.

Our method needs to be validated on more varied data cohorts and including severe structural airway abnormalities due to lung diseases, such as cystic fibrosis or COPD, as well as other lung abnormalities not present in the training datasets. In Chapter 2 we showed that the method had good generalizability and performed well on the highly heterogeneous EXACT’09 dataset, which includes 1) 65% of CT scans with different lung and airway abnormalities, often due to bronchiectasis, 2) different scanning protocols, 3) two expiratory CT scans and 4) three CT angiography (CTA) scans. A more complex and varied dataset where to evaluate our method is the cohort from Chapter 8 of patients with severe non-cystic fibrosis bronchiectasis. In preliminary work not included in this thesis, we tested our method on this cohort and observed that the resulting airway segmentations included more peripheral branches than the segmentations based on manual centerlines that were used in Chapter 8, suggesting a possible advantage of automated analysis. However, the accuracy of our method on this dataset needs to be further studied. Additionally, our method should be evaluated on 1) CT scans with different scanning protocols than those in the EXACT’09 dataset, or 2) more CT scans with some of the protocols included in EXACT’09, particularly for expiratory and CTA scans. Furthermore, our method can be tested on data cohorts with airway abnormalities due to other lung diseases not considered in this thesis, such as asthma, pneumonia or COVID.

2. The proposed deep learning methods, 1) the label refinement method using synthetic labels with errors, and 2) the combination of the U-Net

with graph neural networks, are promising to improve regular segmentation networks.

In Chapter 3 we showed that our label refinement method can correct errors and improve the accuracy of the segmentations obtained by the U-Net. For airway segmentation, by correcting relevant false negative errors, the completeness of our method in Chapter 2 was significantly increased. This is highly beneficial as state-of-the-art methods for airway segmentation have a limited completeness, as we showed in Chapter 2 in the evaluation on the public EXACT'09 dataset, where the best performing methods missed roughly 30% of the total length of the airway tree annotations. Moreover, in Chapter 3 we showed that training the proposed method with semi-supervised learning techniques and additional unlabeled data can improve its performance when compared to supervised learning. This is highly relevant for airway segmentation, as a large amount of CT data do not have ground truth segmentations and otherwise cannot be used.

A limitation of our label refinement method is that it requires the design of synthetic errors specific for the segmented structure, which is a complex image processing task. In Chapter 3 we generated synthetic errors for airway segmentation limited to two types of false negative errors, yet relevant for general cases. When applying the method on CT scans with different structural airway abnormalities, such as those in Chapters 7 and 8, new types of errors and possibly more complex will likely be present, thus requiring additional work to design new synthetic errors.

In Chapter 4 we showed that the combination of the U-Net with graph neural networks can slightly increase the completeness of our method in Chapter 2 for airway segmentation, at the expense of more false positive errors. Graph neural networks are by design suited to analyze tree-like structures, and its combination with the U-Net in an end-to-end optimized approach is promising, so that the U-Net can extract powerful local features to be used by the graph model. However, our method did not show a clear gain in performance when compared to the U-Net. This may be due to the proposed design using a small graph model in the deepest level of the U-Net. Moreover, in our method we had to impose several constraints to the computation of the graph adjacency due to its high runtime and memory footprint. We believe that a method design using a larger and constraint-free graph model that replaces the skip connections in an upper level of the U-Net, or ideally several skip connections, could have achieved a significantly better performance than our method tested in Chapter 4. Additionally, an ablation study to quantify the benefit of including the graph model in the U-Net is needed.

3. The proposed tools that complement the airway segmentation method, 1) to create ground truth segmentations, and 2) to segment the lumen and wall surfaces of airways, are promising and make the method more easily applicable to other cohorts.

A limitation of our method in Chapter 2 is that it requires ground truth airway segmentations for training, as is in general the case for learning-based segmentation methods. Airway references are complex and tedious to obtain manually, requiring roughly 15 hours per CT scan [17], and their completeness is often limited due to small branches being missed by the annotators. In the two training datasets we used in

Chapter 2, the ground truth segmentations were obtained by refining airway references that were either 1) manual annotations of the airway centerlines, which were limited in number, or 2) automated airway extractions, which were of limited quality. To alleviate this limitation, in Chapter 5 we showed a promising approach to reduce the time and effort needed to generate ground truth airway segmentations for model training. This approach is simple to perform and uses open-source software tools readily available: our software in the public repository provided in Chapter 2, and 3D Slicer [123]. Thus, it can be implemented as an easy annotation protocol for future datasets.

In Chapter 6 we showed that our pipeline, combining two existing open-source methods, is reliable and efficient in obtaining both the airway lumen and wall segmentations. It can be applied to new cohorts with a relatively small manpower investment, needing only 1) to retrain the method in Chapter 2 on a small subset of 20–30 CT scans from the cohort, with annotations that can be obtained with our method in Chapter 5, and 2) to adapt parameters of the optimal-surface graph-cut method to the given CT protocol, for instance using an airway phantom. This pipeline can be used to automatically obtain the airway measurements needed to compute quantitative airway biomarkers, such as those in Chapters 7 and 8.

4. The automated airway biomarkers, 1) the median luminal intra-branch tapering over all the airways, and 2) the median airway-artery ratio and airway tapering over the small airways, are promising to quantify structural lung abnormalities from CT scans.

In Chapter 8 we showed that luminal intra-branch tapering measures can be used to quantify severe bronchiectasis, as a reduced median intra-branch tapering correlated well with the BEST-CT scores. In Chapter 7, different airway-artery ratio and airway tapering measures (including the luminal intra-branch tapering) were not sensitive enough to detect improvements in structural cystic fibrosis lung disease due to CFTR therapy. This could be due to the small and retrospective cohort used in this longitudinal study, whose outcome is affected by the use of non-standardized intervals between CT scans. These automated airway measures have been used successfully in other studies to quantify bronchiectasis and airway wall thickening due to cystic fibrosis lung disease, with airway-artery ratio [62, 70, 71] to a greater extent than airway tapering [65, 200]. Automated airway biomarkers are promising to quantify structural abnormalities due to lung diseases. In contrast to visual scoring systems, they can provide objective and reproducible scores, and can be obtained effortlessly. Computing the airway tapering is advantageous over the airway-artery ratio as 1) its evaluation does not require the dimensions of the adjacent arteries, and 2) airway tapering measures are less influenced by differences in the lung volumes and scanning parameters [65].

A limitation of the airway biomarkers studied in Chapters 7 and 8 is that we used the median as the descriptive score of the distribution of airway measures extracted per subject. While the median is robust as it is independent of data outliers that can be due to inaccurate measurements, it also neglects other relevant extreme data measurements. Consequently, the median may be not sensitive enough to capture lung diseases that are heterogeneous and characterized by localized structural airway

abnormalities. This is the case for both cystic fibrosis, where the disease starts in the smaller airways [110] and its severity can vary widely between areas in the lungs [201], as well as bronchiectasis, which can be diffuse or localized depending on its underlying causes [202]. Other descriptive scores that can be more sensitive to capture localized bronchiectasis, while still neglecting extreme outliers, are larger percentiles (e.g., 90% or 95% percentiles).

Another limitation of the automated airway biomarkers in Chapters 7 and 8 is that we used airway segmentations obtained with an automated method that segments airways around manual annotations of the airway centerlines. Several small branches were missed by the annotators in various CT scans, especially for the cohort in Chapter 8 with complex airway trees due severe structural airway deformations. Consequently, the computed biomarkers did not quantify these small branches and this may have reduced their sensitivity, especially for those used in Chapter 7 which were evaluated over the smaller airways. An alternative could have been to use the method in Chapter 2 to automatically detect airways, and use these either to replace or to extend the manual centerline annotations. However, this method was not sufficiently mature at the time of performing the studies in Chapters 7 and 8, especially for the latter with a very complex data cohort.

9.2 Implications for clinical practice

The methods developed in this thesis can be integrated into medical imaging software to improve existing clinical workflows in radiology. We have validated these methods in a few data cohorts and in a research setting, however for them to be used in clinical practice they need to meet more strict requirements for validation. The airway segmentation method in Chapter 2 is especially promising as it performed well on various datasets and is fully automatic and easy to operate, requiring few to no user-defined input settings. The open-source software in the public repository provided in Chapter 2 could be a good starting point to implement this method in medical software.

At the current time, the use of artificial intelligence (AI) methods for clinical practice is scarce and often limited to large academic centers [203]. To reach a widespread implementation of AI methods into clinical routine, some relevant aspects to tackle are:

1. **Training datasets:** image data cohorts used to train AI algorithms are often from the same medical center, with a limited range of population characteristics, scanner parameters and protocols. Additionally, this can lead to unrecognized biases in the training data [204]. Varied training datasets are required to build robust AI methods for clinical routine, including diverse imaging data from different regions and acquired with scanners from different vendors. This makes crucial the sharing of medical data between different medical centers. However, this is restricted due to privacy concerns, among others. An alternative to alleviate this limitation is federated learning [205], where the AI algorithm is trained at different locations while the data remains in the medical center.

2. **Testing datasets:** most AI algorithms reported in research are trained, validated, and tested in a research setting, isolated from clinical practice [206], and few methods include validation on independent external data. Testing datasets must be representative, yet independent, and of very high quality as they are used to evaluate the performance of AI methods and obtain regulatory approval. Moreover, testing datasets are typically from retrospective studies, which do not represent a routine clinical setting. Prospective multi-center trials are needed to evaluate the performance of AI methods for clinical diagnosis [206].
3. **Ground truth:** the availability of image annotations that can be used as ground truth to train, validate, and test AI algorithms is an important bottleneck for manufacturing AI medical software [207]. Annotations are tedious and need to be obtained by clinical experts or medical students. An alternative to alleviate this limitation is crowd-funding, where cost-effective large-scale annotations can be obtained by the general non-expert population [208, 209]. A previous study on liver segmentation from CT data [208] showed that annotations from non-experts, medical students and radiologists had similar accuracy, while the latter were more time efficient.
4. **Technical deployment:** technical integration of AI medical software into the IT infrastructure of medical centers is the leading bottleneck for their implementation into the clinical routine [210]. Computational resources required by AI software can rapidly exceed the capabilities of local servers. A cost-effective solution for this is cloud computing, but this poses several challenges regarding security and data access privileges.
5. **Security:** AI algorithms are often regarded as “black-boxes”, and security and safety is crucial for their implementation in clinical routine. Adversarial attacks on AI methods for imaging data are particularly relevant [211, 212]. These attacks are small imperceptible perturbations to the image data that are designed to intentionally mislead the method into giving the wrong predictions. Several defense methods against adversarial attacks have been proposed [213], however they are only effective against some types of attacks [212] and universal protection is currently not possible.

Any medical software to be used in clinical practice must be approved by regulatory bodies such as the American Food and Drug Administration (FDA) and the European Medicine Agency (EMA), which impose strict requirements for software validation [214, 215]. For AI methods, this regulatory framework is not ideal [215] and modifications are being currently proposed [214]. Particularly, retraining AI methods with new datasets is generally considered as software modification and requires new FDA clearance. At the current time, roughly 160 AI methods in medical imaging software have received approval by the FDA [216]. A relevant example for the research in this thesis is the lung quantification software LungQ (Thirona, Nijmegen, the Netherlands). These cleared AI software tools assist radiologists with their workload of repetitive tasks, enabling the automatic extraction of information and automated analysis, while experts can be in charge of the decision making and diagnosis. Radiologists should also receive training in AI to facilitate their acceptance of AI methods in clinical routine [217].

9.3 Recommendations for future research

A list of recommendations and possible directions to continue the research presented in this thesis is given below:

1. To further improve the developed airway segmentation methods.

Our method in Chapter 2 needs to be validated in more data cohorts. For every new dataset, the method can be retrained by reserving a small subset of the CT data to be added to the existing training datasets. To obtain the ground truth airway segmentations needed for training, we recommend using the proposed approach in Chapter 5 to correct initial segmentations obtained with the best pre-trained U-Net model available. These steps can be implemented as a protocol to gradually improve the generalizability of our method.

Some improvements to the pipeline for our method in Chapter 2 are straightforward. First, the U-Net can be used to automatically obtain the masks of the lungs and the trachea and main bronchi needed as inputs to the method, contrary to using the region growing method which requires user-defined settings. Second, image elastic deformations [56] can be implemented as data augmentation. This can improve the performance of our method for a limited overhead in runtime during training [77].

Our method in Chapter 2 can be extended to segment both the airway lumen and the wall, using a U-Net backbone for dual-class segmentation, so that both surfaces are obtained in an end-to-end optimized setting. Additionally, this could improve the accuracy of the lumen segmentation as the U-Net can learn more specialized features that discriminate between the lumen and wall surfaces. The ground truth airway wall surfaces needed for training can be obtained for instance with the optimal-surface graph-cut method [58] used in Chapters 6, 7 and 8. Moreover, the U-Net in our method can be used to obtain the segmentation of the vascular tree in the lungs [52, 218, 219]. The resulting segmentations will likely be more accurate than those used in Chapter 7 to compute the airway-artery ratio. The ground truth vessel segmentation needed for training can be obtained with previous methods [32].

An idea to improve our method in Chapter 2 is to use the prior structural knowledge of airways as being adjacent to an artery [16, 32]. This information can be incorporated in the U-Net by adding a penalty term to the training loss that penalizes the prediction of airways distant from arteries. This penalty can be formulated for instance as the L^2 -norm of the product of the airway probability maps output by the network and a voxelwise distance map to the closest artery. Computing the distance map requires the segmentation of the vascular tree in the lungs, which can be obtained with the U-Net [52, 218, 219].

Semi-supervised and weakly-supervised learning techniques [220] may be useful for airway segmentation. With semi-supervised learning, unlabeled CT data can be used during training to extract additional relevant information by the network. In Chapter 3 we showed that using semi-supervised learning to train our label refinement method increased the performance of the method. With weakly-supervised learning, weak labels or annotations that are coarser and cheaper than voxelwise airway segmentations can be used for training. Examples of weak labels can be the PRAGMA-CF and BEST-CT visual scores used in Chapters 7 and 8, respectively, which can give insights

to the network into which lung regions have structural abnormalities and thus detecting airways is more difficult. Finally, transfer learning, and particularly domain adaptation techniques [220], can be useful to obtain more robust airway segmentations across different scanners and scanning protocols.

2. To improve the automated airway biomarkers to quantify lung abnormalities from CT scans.

The airway biomarkers proposed in Chapters 7 and 8 need to be further validated. They can be evaluated on longitudinal studies using prospective cohorts and standardized scanning intervals and protocols. For cystic fibrosis lung disease, the airway tapering has been relatively poorly studied to assess structural airway abnormalities [65, 200], in contrast to the airway-artery ratio [17, 62, 68, 69, 70, 71]. If further studies show that the automated airway tapering is equally or more sensitive than the airway-artery ratio, we recommend using it as it is easier to evaluate. Moreover, the airway tapering can be evaluated to capture the heterogeneity of bronchiectasis and cystic fibrosis, by studying the segregation of airway tapering measures according to the airway generation, similarly to other previous studies using the airway-artery ratio [17, 68]. Finally, the airway tapering and airway-artery ratio biomarkers can be evaluated to assess other lung diseases characterized by bronchiectasis and bronchial wall thickening, such as COPD or asthma.

Machine learning methods can be used in combination with the airway measures studied in Chapters 7 and 8 to obtain better airway biomarkers. A neural network can be developed to predict a meaningful biomarker from the equispaced histograms of the distributions of both airway tapering and airway-artery measures, using a relevant patient outcome as ground truth for model training. Moreover, a convolutional neural network can be developed to extract directly the airway measurements from the CT scan, using manual annotations of individual airway dimensions [17, 62] for model training. This approach could be more robust than using an airway segmentation as an intermediate step to extract the measurements, as we did in Chapters 7 and 8.

9.4 Conclusion

9

In this thesis, we have developed automatic image processing methods to segment the bronchial tree from chest CT scans. These methods are based on deep convolutional neural networks, in particular the U-Net, and include various approaches to improve the performance of regular segmentation networks. We validated these methods on several data cohorts, including challenging CT scans with severe airway abnormalities. We showed that convolutional neural networks are a powerful and robust methodology for airway segmentation. Moreover, we extracted automated airway biomarkers to quantify structural lung abnormalities from CT scans. We studied two airway biomarkers, derived from airway tapering and airway-artery ratio measures, to quantify bronchiectasis. The developed methods in this thesis are promising to be integrated in automated imaging tools for clinicians to assess accurately and efficiently pulmonary diseases affecting the airways.

Summary

In this thesis we developed automatic image processing methods to segment the bronchial tree from chest CT scans and subsequently extracted airway measurements. These automated measurements can be used as quantitative biomarkers reflecting clinically relevant structural abnormalities of the airways. Moreover, we applied these automated biomarkers to analyze CT scans for various lung diseases. With these developed methods, we aim to contribute to the development of automated tools for clinicians to assess accurately and efficiently pulmonary diseases affecting the airways.

The content of this thesis is split into two main research topics: **Chapters 2, 3, 4, 5 and 6** address the development of airway segmentation methods from CT scans, while **Chapters 7 and 8** address the application of CT analysis methods to assess various lung diseases from CT scans.

In **Chapter 2** we developed a fully automatic and end-to-end optimized method to segment the bronchial tree. This method is based on the state-of-the-art deep learning U-Net architecture. The method can process at once large 3D patches extracted from the CT scans, which makes it simple, robust and efficient. We validated this method on three datasets with different population characteristics and scanning protocols, that included subjects with a wide range of airway abnormalities. The results show that this method can obtain high-quality and complete airway segmentations, even on challenging CT scans with severe airway abnormalities. The method also showed good generalizability and was accurate on CT scans of different characteristics to those used for training. Moreover, the method achieved top performance scores when compared to other relevant deep learning airway segmentation methods.

In our work in **Chapter 2** we identified two major limitations when building a deep learning airway segmentation method. First, the segmentation of the smaller peripheral branches in the bronchial tree is challenging, and even state-of-the-art methods miss several terminal airways. Thus, there is still room for improvement. Second, training sets are typically small due to the lack of ground truth airway segmentations. To address the first limitation, in **Chapters 3 and 4** we developed two extensions to the airway segmentation method in **Chapter 2** in order to increase its performance. To address the second limitation, in **Chapter 5** we developed an approach to efficiently create airway segmentations for model training.

In **Chapter 3** we developed a novel method to correct errors in the initial segmentations from a U-Net. This method uses a set of synthetic labels augmented with realistic errors, together with the original image as inputs, and is trained to correct

these errors in the segmentation results. As part of the method, we developed an automated tool to generate synthetic errors on the fly during training. We validated this method for both tasks of 1) airway segmentation from chest CT scans and 2) vessel segmentation from brain CTA scans. For airway segmentation, we generated synthetic errors that resemble those found in the segmentation from our method in **Chapter 2**, which are false negative errors of two types: 1) missing terminal branches and 2) discontinuities in the segmented airways. The results show that this method can correct errors and increase the completeness of the segmentation results from the U-Net.

In **Chapter 4** we developed a novel end-to-end optimized segmentation method that combines the U-Net with a graph neural network model. In this method the normal convolutions in the deepest level of the U-Net are replaced by a series of graph convolutions. This method can extract information from a larger region of the images and incorporate this in the features learned in the U-Net. This can help improve the segmentation decisions by the network, and in turn the results. We validated this method on the task of airway segmentation from chest CT scans. The results show that this method can extract slightly more complete airway segmentations than the U-Net, with slightly more false positive errors.

In **Chapter 5** we proposed an approach to efficiently generate ground truth airway segmentations. This approach consists of manually correcting an initial segmentation of the bronchial tree that is obtained with our method in **Chapter 2**. Correcting segmentations was relatively quick using open-source software and required 2–4 hours per CT scan, in contrast to 15 hours for a fully manual segmentation. We validated this approach by correcting the initial airway segmentations obtained with a model trained on a small dataset, and using these as ground truth to retrain the U-Net, validated in cross-validation. The results from the retrained model were more accurate and complete than the initial segmentations.

In **Chapter 6** we developed an automated pipeline to segment both the lumen and wall surfaces of the bronchial tree. This pipeline combines our method in **Chapter 2** for automatic airway extraction with an optimal-surface graph-cut method to obtain the lumen and wall segmentations around the extracted airways. Then bronchial parameters are automatically measured from the segmentations. We assessed the repeatability of this approach in a dataset with two short-term repeated CT scans per subject, by comparing the bronchial parameters obtained in each scan. The results show that the bronchial measurements were reproducible up to the 6th airway generation.

In **Chapters 7 and 8** we conducted two clinical studies using quantitative analysis of CT scans to assess structural lung abnormalities in patients with lung diseases. In both studies, we extracted automated airway biomarkers that are relevant to assess bronchiectasis: airway tapering, airway-artery ratio and wall-artery ratio. Computing these biomarkers requires the segmentation of the airway lumen and outer wall, which we obtained with an automated method that segments the lumen and wall surfaces around manual annotations of the airway centerlines. Additionally, the airway-artery and wall-artery ratios require the segmentation of the blood vessels in the lungs, which we obtained with an automated method that detects elongated tubular structures with high intensity in the CT scan.

In **Chapter 7** we studied the effect of CFTR modulator therapy on structural lung abnormalities in patients with cystic fibrosis lung disease. To quantify these abnormalities from CT scans, we used three different CT analysis methods: a quantitative method based on automated biomarkers, and two visual scoring systems. The automated biomarkers are the median of airway-artery ratio and airway tapering measures extracted over the small airways, identified as being adjacent to a small artery of diameter lower than 3.08 mm. The two visual systems are: 1) PRAGMA-CF, a quantitative grid-based scoring method based on manual annotations for the presence of common abnormalities in cystic fibrosis; and 2) CF-CT, a semi-quantitative method based on an estimate of the overall extent of abnormalities in each lung lobe. The results show that quantitative methods were more sensitive than semi-quantitative ones to detect improvements in structural lung disease due to therapy.

In **Chapter 8** we proposed a quantitative CT analysis method to assess structural lung abnormalities in patients with non-cystic fibrosis bronchiectasis, named BEST-CT, especially developed for these patients. BEST-CT is a grid-based visual scoring method, where a grid is overlaid on axial CT slices and cells are annotated for the presence of common abnormalities in bronchiectasis, according to a predefined list of possible items and in hierarchical order. We validated the BEST-CT method on a cohort of patients with severe bronchiectasis and chronic bacterial infection. We compared BEST-CT with two other methods: 1) a quantitative method based on automated biomarkers, the median of luminal intra-branch tapering measures extracted over all the airways; and 2) Hartmann, a semi-quantitative method based on an estimate of the overall extent of abnormalities in each lung lobe. The results show that the BEST-CT method can capture the heterogeneity of structural lung changes due to bronchiectasis, and the BEST-CT scores correlate well with the intra-branch tapering and the Hartmann scores.

Samenvatting

In dit proefschrift hebben we automatische beeldverwerking methoden ontwikkeld om de luchtwegboom te segmenteren uit CT-scans van de borstkas en vervolgens luchtwegmetingen te extraheren. Deze geautomatiseerde metingen kunnen worden gebruikt als kwantitatieve biomarkers die klinisch relevante structurele afwijkingen van de luchtwegen weergeven. Binnen dit proefschrift hebben we deze geautomatiseerde biomarkers toegepast om CT-scans te analyseren voor verschillende longziekten. Met deze ontwikkelde methoden willen we bijdragen tot de ontwikkeling van geautomatiseerde hulpmiddelen voor klinici om longaandoeningen die de luchtwegen aantasten nauwkeurig en efficiënt te beoordelen.

De inhoud van dit proefschrift is opgesplitst in twee belangrijke onderzoeksthema's: de **Hoofdstukken 2, 3, 4, 5 en 6** behandelen de ontwikkeling van luchtwegsegmentatiemethoden uit CT-scans, terwijl de **Hoofdstukken 7 en 8** de toepassing van CT-analysemethoden behandelen om verschillende longaandoeningen uit CT-scans te beoordelen.

In **Hoofdstuk 2** hebben we een volledig automatische en end-to-end geoptimaliseerde methode ontwikkeld om de luchtwegboom te segmenteren. Deze methode is gebaseerd op het state-of-the-art deep learning U-Netwerk. De methode kan in één keer grote driedimensionale patches verwerken die uit de CT-scans zijn geëxtraheerd, wat de methode eenvoudig, robuust en efficiënt maakt. We hebben deze methode gevalideerd op drie datasets met verschillende populatiekenmerken en scanprotocollen, die proefpersonen omvatten met een breed scala aan luchtwegafwijkingen. De resultaten tonen aan dat deze methode kwalitatief goede en volledige luchtwegsegmentaties kan verkrijgen, zelfs op uitdagende CT-scans met ernstige luchtwegafwijkingen. De methode bleek ook goed generaliseerbaar en was ook nauwkeurig op CT-scans met andere kenmerken dan die gebruikt werden voor de training. Bovendien behaalde de methode topprestaties in vergelijking met andere relevante deep learning luchtwegsegmentatiemethoden.

In ons werk in **Hoofdstuk 2** hebben we twee belangrijke beperkingen geïdentificeerd bij het bouwen van een deep learning luchtwegsegmentatiemethode. Ten eerste, de segmentatie van de kleinere perifere luchtwegen is een uitdaging, en zelfs state-of-the-art methoden missen een aantal perifere luchtwegen. Er is dus nog ruimte voor verbetering. Ten tweede, training sets zijn meestal klein. Om de eerste beperking aan te pakken, hebben we in **Hoofdstukken 3 en 4** twee uitbreidingen ontwikkeld op de luchtwegsegmentatiemethode uit **Hoofdstuk 2** om de prestaties te verbeteren. Om de tweede beperking aan te pakken, hebben we in **Hoofdstuk 5** een aanpak

ontwikkeld om efficiënt meer luchtwegsegmentaties te maken voor het trainen van verbeterde modellen.

In **Hoofdstuk 3** hebben we een nieuwe methode ontwikkeld om fouten in de initiële segmentaties van een U-Net te corrigeren. Deze methode gebruikt een set van synthetische labels, aangevuld met realistische fouten, samen met het originele beeld als input, en wordt getraind om deze fouten in de segmentatie resultaten te corrigeren. Als onderdeel van de methode hebben we een geautomatiseerd instrument ontwikkeld om synthetische fouten te genereren tijdens de training. We hebben deze methode gevalideerd voor beide taken van 1) luchtwegsegmentatie in CT-scans van de borstkas en 2) vaatsegmentatie in CT-scans van de hersenen. Voor luchtwegsegmentatie hebben we synthetische fouten gegenereerd die lijken op de fouten die gevonden zijn in de segmentatie van onze methode uit **Hoofdstuk 2**, namelijk vals-negatieve fouten van twee typen: 1) ontbrekende eindtakken en 2) discontinuïteiten in de gesegmenteerde luchtwegen. De resultaten laten zien dat deze methode fouten kan corrigeren en de volledigheid van de segmentatie resultaten van het U-Net kan verhogen.

In **Hoofdstuk 4** hebben wij een nieuwe end-to-end geoptimaliseerde segmentatiemethode ontwikkeld die het U-Net combineert met een graaf neuraal netwerk model. In dit model worden de normale convoluties in het diepste niveau van het U-Net vervangen door een serie van graaf convoluties. Zo kan de methode informatie extraheren uit een groter gebied en deze opnemen in de kenmerken die in het U-Net worden geleerd. Dit kan helpen om de door het netwerk genomen beslissingen te verbeteren, en op zijn beurt de resultaten. We hebben deze methode gevalideerd op de taak van luchtwegsegmentatie uit CT-scans van de borstkas. De resultaten tonen aan dat deze methode iets vollediger luchtwegsegmentaties kan maken dan het U-Net, met iets meer vals-positieve fouten.

In **Hoofdstuk 5** hebben we een aanpak voorgesteld om efficiënt annotaties van luchtwegen te genereren. Deze aanpak bestaat uit het handmatig corrigeren van een initiële segmentatie van de luchtwegboom die is verkregen met onze methode in **Hoofdstuk 2**. Het corrigeren van segmentaties was relatief snel met behulp van open-source software, en vereiste 2–4 uur per CT scan, vergeleken met 15 uur voor een volledig handmatige segmentatie. We hebben deze aanpak gevalideerd door de initiële luchtwegsegmentaties behaald met een model getraind op een kleine dataset te corrigeren, en deze te gebruiken om het netwerk opnieuw hertrainen, gevalideerd in kruisvalidatie. De resultaten van het opnieuw getrainde model bleken nauwkeuriger en vollediger te zijn dan de aanvankelijke segmentaties.

In **Hoofdstuk 6** hebben we een volledige, geautomatiseerde pijplijn ontwikkeld om zowel het lumen als de wand van de luchtwegboom te segmenteren. Deze pijplijn combineert onze methode in **Hoofdstuk 2** voor automatische luchtweg extractie met een optimal-surface graph-cut methode om de lumen- en wandsegmentaties rond de geëxtraheerde luchtwegen te verkrijgen. Vervolgens worden bronchiale parameters automatisch gemeten op basis van de segmentaties. Wij hebben deze aanpak gevalideerd in een dataset met twee binnen een korte periode herhaalde CT-scans per proefpersoon. De resultaten tonen aan dat de metingen reproduceerbaar zijn tot en met de 6^{de} luchtweggeneratie.

In de **Hoofdstukken 7 en 8** hebben wij twee klinische studies uitgevoerd met gebruikmaking van kwantitatieve analyse van CT-scans om structurele longafwijkingen

bij patiënten met longaandoeningen te beoordelen. In beide studies hebben we geautomatiseerde luchtwegbiomarkers geëxtraheerd die relevant zijn voor de beoordeling van bronchiëctasieën: verandering in luchtwegdiameter, luchtweg-arterie verhouding en luchtwegwand-arterie verhouding. Berekening van deze biomarkers vereist de segmentatie van het lumen en de buitenwand van de luchtwegen, die we verkregen met een geautomatiseerde methode die het lumen en de wand segmenteert rond handmatige annotaties van de luchtwegmiddellijnen. Bovendien vereisen de luchtweg-arterie en luchtwegwand-arterie verhoudingen de segmentatie van de bloedvaten, die we verkregen met een geautomatiseerde methode die langgerekte cilindervormige structuren met hoge intensiteit in de CT scan detecteert.

In **Hoofdstuk 7** bestudeerden we het effect van een therapie op structurele longafwijkingen bij patiënten met cystische fibrose longziekte. Om deze afwijkingen op basis van CT-scans te kwantificeren, gebruikten we drie verschillende CT-analyse methoden: een kwantitatieve methode op basis van geautomatiseerde biomarkers, en twee visuele scoring systemen. De geautomatiseerde biomarkers zijn de mediaan van de luchtweg-arterie verhouding en de verandering in luchtwegdiameter die geëxtraheerd worden over de kleine luchtwegen, geïdentificeerd als grenzend aan een kleine slagader met een diameter kleiner dan 3.08 mm. De twee visuele systemen zijn: 1) PRAGMA-CF, een kwantitatieve raster-gebaseerde scoringsmethode op basis van handmatige annotaties voor de aanwezigheid van veel voorkomende afwijkingen bij cystische fibrose; en 2) CF-CT, een semikwantitatieve methode op basis van een schatting van de totale omvang van afwijkingen in elke longkwab. De resultaten tonen aan dat kwantitatieve methoden gevoeliger waren dan semikwantitatieve methoden om verbeteringen in structurele longaandoeningen als gevolg van therapie te detecteren.

In **Hoofdstuk 8** hebben we een kwantitatieve CT-analyse methode voorgesteld om structurele longafwijkingen te beoordelen bij patiënten met niet-cystische fibrose bronchiëctasieën, BEST-CT genaamd, speciaal ontwikkeld voor deze patiënten. BEST-CT is een raster-gebaseerde visuele scoringsmethode, waarbij een raster over de axiale CT doorsneden wordt gelegd en cellen worden geannoteerd voor de aanwezigheid van veel voorkomende afwijkingen in bronchiëctasieën, volgens een lijst van mogelijke items en in hiërarchische volgorde. We hebben de BEST-CT methode gevalideerd op een cohort van patiënten met ernstige bronchiëctasieën en chronische bacteriële infectie. We vergeleken BEST-CT met twee andere methoden: 1) een kwantitatieve methode op basis van geautomatiseerde biomarkers, de mediaan van de verandering in lumenale luchtwegdiameter metingen geëxtraheerd over alle luchtwegen; en 2) Hartmann, een semikwantitatieve methode op basis van een schatting van voor de totale omvang van afwijkingen in elke longkwab. De resultaten tonen aan dat de BEST-CT methode de heterogeniteit van structurele long veranderingen ten gevolge van bronchiëctasieën kan vastleggen, en dat de BEST-CT score goed correleert met de andere methoden.

Acknowledgments

This is how my PhD started: I had left a secure job in the aerospace industry to learn from scratch a new field, machine learning, and in a new country. Looking back, it has been an enriching PhD journey of nearly 5 years, full of lessons and good moments, but also plenty of struggles and moments of doubts on whether I had made the right decision. And I am lucky to have shared this time with amazing people that I can now call friends. Since the start of my PhD I never allowed myself to picture the moment of submitting my PhD thesis and writing the acknowledgments section. Nevertheless here I am, and here it goes:

Firstly, to my PhD supervisors:

Marleen, I am very thankful for your support and guidance throughout my PhD, and for believing that I could quickly learn the medical imaging field at the start. I learned a lot from your technical expertise, research integrity and high work standards. Your detailed oversight of my work, critical feedback and constant drive to do better, especially during moments of low motivation, was crucial for the success of this PhD. Thank you for all your comments while reviewing my papers and thesis.

Harm, thank you for mentoring me in lung analysis, guiding my work towards more clinical impact, and providing me with critical feedback throughout my PhD. A big takeaway of my collaboration with you was how “taking full ownership is critical for success”. This mindset was crucial to finalize my PhD, and empowers me now in my professional career and other life projects.

Secondly, to my work colleagues and now friends:

Ivan, I really enjoyed working together, and our collaboration made my PhD more fun. We complement each other well, and I hope we can collaborate again in the future. Thank you for your patience when answering my clinical questions. **Raghav**, I am thankful that we worked together at the start of my PhD. I learned a lot about airway analysis and graph methods from you, as well as being more critical about our work. **Martijn**, slowly but surely we developed a good friendship. I enjoyed our discussions around a coffee in the early mornings. Your high work discipline while keeping a good work-life balance motivated me a lot during my PhD. **Robin**, I am happy that we became friends quickly, by sharing similar interests, several languages to communicate and our determination to learn Dutch. You were always keen on showing me relevant technical content. **Alex**, I enjoyed working together, despite the challenges that we faced in our project. Your determination to finish your PhD while holding a full-time job motivated me to do the same.

To the BIGH group members: **Jose**, your drive to push through your PhD despite the challenges you faced was inspiring. Thank you for helping me find my way around BIGH at the start. **Kim**, your constant happiness and positive energy was a real booster in both good and bad moments. I enjoyed our long discussions across our desks. **Shuai**, I liked working together and got to know a lot from you about deep learning. Thank you for coordinating the submission of our joint paper after I started a full-time job. **Gijs**, I am thankful for the many times you helped me in great detail both in my projects and other non-work related matters. **Zahra, Gerda, Florian, Annegreet** and **Deep**, thank you for your critical feedback of my work during our group meetings. **Sebastian, Wietske, Karin, Arno, Gokhan, Gennady, Jiahang, Bo** and **Laura**, thank you for the nice conversations and moments we shared together.

To the LA group members: **Jennifer**, I enjoyed working together and got to know a lot from you about airway diseases. Thank you for helping me find my way around the LA group at the start. **Mariette**, thank you for coordinating our contributions to the iABC project. **Qianting, Federico, Merlijn, Wytse, Pierluigi** and **Bernadette**, thank you for the nice conversations and moments we shared together.

Thirdly, to my friends and my family:

Andrei, thank you for encouraging me to continue and enjoy my PhD when I had plenty of doubts. I always enjoy our long debates about life and politics. **Fran**, I learned a lot from your positive mindset and resilience in the face of challenges. We both can prove that patience is key to achieve good long-term outcomes. **Timur** and **Elena, Antonio** and **Sophie**, thank you for your good and sincere advice throughout my PhD. I enjoyed meeting for dinner and the nice trips we did together. Especially thank you for witnessing one of the happiest moments of my life: my wedding. **Adrian**, I appreciate the times we see each other and catch up, although less often now due to COVID. **Felipe, Hannes, Ryan, Anastasia** and **Eelco**, I am happy that we kept our friendship intact despite moving far away from each other. I really enjoy having our regular group chats.

Mis padres **Antonio** y **María del Carmen**, vosotros me habéis dado la educación e inculcado los valores de trabajo, disciplina y humildad que han hecho posible completar mi doctorado. Gracias por vuestro apoyo y confianza continuos, sobre todo al tomar decisiones arriesgadas. Mis hermanos **María** y **Daniel**, gracias por apoyarme desde la distancia y por desearme lo mejor. Sabéis que estaré siempre allí a vuestro lado para lo que necesitéis.

Ma belle famille **Daniel, Chantal, Joanna** et **Dustin**, vous m'avez réconforté dans les moments où mon doctorat me sortait par les yeux. Merci d'avoir réussi à me faire rire dans ces moments-là.

And lastly, but most importantly, **Mathilde**, my wife, life partner, companion of adventures and mother of my upcoming son. Meeting you was the most important achievement of my PhD time. You supported me through many difficult moments, and put up with me being a workaholic during the past two years. You listened to me complaining about my PhD and doubting whether I should quit, and yet encouraged me and kicked my ass to continue and finish it. You helped me become more organized, result-driven and a better communicator. You were crucial for the completion of this PhD. From the bottom of my heart, I am infinitely thankful to you.

About the author

Antonio García-Uceda Juárez was born in Madrid, Spain on the 7th of August 1989. He attended primary and secondary education at the school “Los Olivos”, in Mejo-rada del Campo (Madrid), Spain. Then, he started a degree in Mechanical Engineering at the “Escuela Técnica Superior de Ingenieros Industriales” at the Polytechnic University of Madrid. During the last year of his studies he enrolled in a master’s degree in Computational Fluid Dynamics (CFD) at Cranfield University, UK, as part of a double degree program. In his MSc thesis, he implemented a turbulence model in a numerical simulation code for aerodynamics, and applied it to investigate the interaction between shock waves and boundary layers. He graduated as a Mechanical Engineer at the top of his class in 2012, with a master’s degree in CFD.



In October 2012, Antonio moved to Brussels, Belgium, to start working as a CFD software engineer at NUMECA International. During five years he developed software based on novel high-order finite element methods for numerical simulations of fluid dynamics. He also parallelized this software to be run in massive parallel supercomputers (thousands of cores). During this time he participated in various EU consortia and collaborated with several research institutions.

In November 2017, Antonio moved to Rotterdam, The Netherlands, to start a PhD program at the Biomedical Imaging Group Rotterdam (BGR), at Erasmus Medical Center (MC). During the four years of this PhD he developed novel image processing methods for the automatic analysis of chest CT scans and assessment of lung diseases. He also attended several international scientific conferences in medical image computing. A significant part of these four years are summarized in this thesis.

In September 2021, Antonio started working as a project lead in artificial intelligence at Dümme Orange. In this new role he is developing and deploying image analysis and data science methods to assist breeding of ornamental plants.

Publications

Journal Papers

A. Garcia-Uceda, R. Selvan, Z. Saghir, H. Tiddens, and M. de Bruijne, “Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks,” *Scientific Reports*, vol. 11, no. 1, p. 16 001, 2021. DOI: 10.1038/s41598-021-95364-1.

I. Dudurych, **A. Garcia-Uceda**, Z. Saghir, H. Tiddens, R. Vliegenthart, and M. de Bruijne, “Creating a training set for artificial intelligence from initial segmentations of airways,” *European Radiology Experimental*, vol. 5, no. 1, p. 54, 2021. DOI: 10.1186/s41747-021-00247-9.

S. Chen^{*}, **A. Garcia-Uceda**^{*}, J. Su^{*}, G. van Tulder, L. Wolff, T. van Walsum, and M. de Bruijne, “Label refinement network from synthetic error augmentation for medical image segmentation,” *Submitted*.

I. Dudurych^{*}, **A. Garcia-Uceda**^{*}, J. Petersen, Y. Du, R. Vliegenthart, and M. de Bruijne, “Reproducibility of a combined artificial intelligence and optimal-surface graph-cut method to automate bronchial parameter extraction,” *Submitted*.

L. Mok^{*}, **A. Garcia-Uceda**^{*}, M. Cooper, M. Kemner van de Corput, M. de Bruijne, N. Feyaerts, T. Rosenow, K. De Boeck, and S. Stick, “The effect of CFTR modulators on structural lung disease in cystic fibrosis,” *Submitted*.

J. Meerburg, **A. Garcia-Uceda**, O. Dragt, E. Andrinopoulou, M. Kemner van de Corput, P. Ciet, G. Angyalosi, J. Elborn, J. Chalmers, M. Tunney, M. de Bruijne, H. Tiddens, and on behalf of the iABC study group, “Quantitative chest computed tomography scoring technique for bronchiectasis (BEST-CT),” *Submitted*.

A. Cristobal-Huerta^{*}, **A. Garcia-Uceda**^{*}, E. Oei, D. Poot, and J. Hernandez-Tamames, “Enhancing high-resolution 3D-GRASE knee imaging by deep convolutional neural networks,” *Submitted*.

R. Selvan, T. Kipf, M. Welling, **A. Garcia-Uceda Juarez**, J. Pedersen, J. Petersen, and M. de Bruijne, “Graph refinement based airway extraction using mean-field networks and graph neural networks,” *Medical Image Analysis*, vol. 64, p. 101751, 2020. DOI: <https://doi.org/10.1016/j.media.2020.101751>.

Conference Papers

A. Garcia-Uceda Juarez, R. Selvan, Z. Saghir, and M. de Bruijne, “A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs,” in *Machine Learning in Medical Imaging*, 2019, pp. 583–591. DOI: 10.1007/978-3-030-32692-0_67.

A. Garcia-Uceda Juarez, H. Tiddens, and M. de Bruijne, “Automatic airway segmentation in chest CT using convolutional neural networks,” in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, 2018, pp. 238–250. DOI: 10.1007/978-3-030-00946-5_24.

S. Chen, G. Bortsova, **A. Garcia-Uceda Juarez**, G. van Tulder, and M. de Bruijne, “Multi-task attention-based semi-supervised learning for medical image segmentation,” in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2019*, 2019, pp. 457–465. DOI: 10.1007/978-3-030-32248-9_51.

Presentations

A. Garcia-Uceda Juarez[†], R. Selvan, Z. Saghir, and M. de Bruijne, “A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs,” Presented at MICCAI 2019, Shenzhen, China, Oct. 2019.

A. Garcia-Uceda Juarez[†], H. Tiddens, and M. de Bruijne, “Automatic airway segmentation in chest CT using convolutional neural networks,” Presented at MICCAI 2018, Granada, Spain, Sep. 2018.

* indicates equal contributions

† indicates presenting author

PhD portfolio

Courses	Year	ECTS
Front-End Vision and Multi-Scale Image Analysis <i>Eindhoven University of Technology, The Netherlands</i>	2017	4
Advanced Pattern Recognition <i>Delft University of Technology, The Netherlands</i>	2018	2
Medical Imaging Summer School <i>University of Catania, Italy</i>	2018	4
A Programmer's Guide for Modern High Performance Computing <i>University of Amsterdam, The Netherlands</i>	2018	3
Bayesian statistics and JASP <i>Erasmus MC, The Netherlands</i>	2018	0.6
Computer Vision by Learning <i>University of Amsterdam, The Netherlands</i>	2019	4
Good Practices for High Performance Computing and Cloud <i>SURFsara computing services, The Netherlands</i>	2019	1
Data, lights, camera, action! Scientific visualization done beautifully using Blender <i>SURFsara computing services, The Netherlands</i>	2019	0.5
Biostatistical Methods I: Basic Principles <i>Erasmus MC, The Netherlands</i>	2019	5.7
Scientific Integrity <i>Erasmus MC, The Netherlands</i>	2019	0.3
Total		25.1

Online Courses	Year	ECTS
Machine Learning <i>Stanford University, USA</i>	2018	2
Writing in the Sciences <i>Stanford University, USA</i>	2020	3
Code Refinery workshop <i>eScience center, The Netherlands</i>	2021	1
Total		6

International and local research meetings	Year	ECTS
NVPHBV <i>Eindhoven, The Netherlands</i>	2018	0.3
The Medical Image Computing and Computer Assisted Intervention Society <i>Granada, Spain</i>	2018	2
The Medical Image Computing and Computer Assisted Intervention Society <i>Shenzhen, China</i>	2019	2
Biomedical imaging group seminars (biweekly) <i>Erasmus MC, The Netherlands</i>	2017 – 2021	1
Medical Informatics research meetings (biweekly) <i>Erasmus MC, The Netherlands</i>	2017 – 2021	1
Total		6.3

Grants & Awards	Year
NVidia GPU grant (together with Dr. Marleen de Bruijne)	2018

Acronyms

2D Two Dimensional.

3D Three Dimensional.

4D Four Dimensional.

AA Airway-Artery.

AAR Airway-Artery Ratio.

AD Airway Disease.

AdaBoost Adaptive Boosting.

AI Artificial Intelligence.

ANCOVA ANalysis of COVariance.

ANOVA ANalysis Of VAriance.

ATCON ATelectasis / CONsolidation.

AWT Airway Wall Thickening.

BE BronchiEctasis.

BEMP BronchiEctasis with Mucus Plugging.

BEST-CF BronchiEctasis Scoring Technique for CT.

BEwMP BronchiEctasis without Mucus Plugging.

BMI Body Mass Index.

BSI Bronchiectasis Severity Index.

CF Cystic Fibrosis.

CFTR Cystic Fibrosis Transmembrane conductance Regulator.

CFU Colony Forming Units.

CI Confidence Interval.

CL Centerline Leakage.

cm centimeter.

CNN Convolutional Neural Network.

COPD Chronic Obstructive Pulmonary Disease.

CPU Central Processing Unit.

CT Computed Tomography.

CTA Computed Tomography Angiography.

CUDA Compute Unified Device Architecture.

CVID Common Variable Immune Deficiency.

dB decibel.

DIS Total (airway) DISEase.

DL Deep Learning.

DLCST Danish Lung Cancer Screening Trial.

DLP Dose Length Product.

DSC Dice Similarity Coefficient.

EMA European Medicine Agency.

EMC Erasmus MC, University Medical Centre Rotterdam.

EMPBUL EMPhysema or BULlae.

EXACT EXtraction of Airways from CT.

FDA Food and Drug Administration.

FEF Forced Expiratory Flow.

FEV Forced Expiratory Volume.

FN False Negative.

FOV Field Of View.

FP False Positive.

FPR False Positive Rate.

FRC Functional Residual Capacity.

FVC Forced Vital Capacity.

FWHM Full-Width at Half-Maximum method.

g gram.

GB GigaByte.

GGO Ground-Glass Opacities.

GLI Global Lung Initiative.

GLILD Granulomatous Lymphocytic Interstitial Lung Disease.

GNN Graph Neural Network.

GPU Graphics Processing Unit.

HA Healthy Airways.

HM Hartmann method.

HP Healthy Parenchyma.

HU Hounsfield Units.

ICC Intra-Class Correlation coefficient.

ImaLife Imaging in Lifelines.

Inter-BT Inter-Branch Tapering.

Intra-BT Intra-Branch Tapering.

IQR InterQuartile Range.

kg kilogram.

kHz kiloHertz.

kNN k-Nearest Neighbors.

kVp kiloVoltage peak.

L liter.

LA Luminal Area.

LAR Low Attenuation Regions.

LASN Label Appearance Simulation Network.

LoA Limits of Agreement.

LR Label Refinement.

m meter.

mAs milliAmpere second.

MDCT MultiDetector Computed Tomography.

mg milligram.

mGycm milliGray per centimeter.

min minute.

ML Machine Learning.

mm millimeter.

MP Mucus Plugging.

MR Magnetic Resonance.

MRA Magnetic Resonance Angiography.

MRI Magnetic Resonance Imaging.

NN Neural Network.

NOR Normal Or high attenuation Regions.

Pa Pseudomonas aeruginosa.

Pi Internal Perimeter.

Pi10 Square root of the wall area of a hypothetical airway with internal perimeter of 10mm.

PRAGMA-CF Perth Rotterdam Annotated Grid Morphometric Analysis for CF.

PwCF People with Cystic Fibrosis.

QOL Quality Of Life.

R2 Coefficient of Determination.

RAM Random-Access Memory.

ReLU Rectified Linear Unit.

ROC Receiver Operating Characteristic.

ROI Region Of Interest.

s second.

SD Standard Deviation.

SGD Stochastic Gradient Descent.

SRWA Square Root of the Wall Area.

TAC Total Airway Count.

TBE Total BronchiEctasis.

TL Tree Length.

TLV Total Lung Volume.

TMP Total Mucus Plugging.

TOF Time-Of-Flight.

TPE Tree-structured Parzen Estimator algorithm.

TPR True Positive Rate.

WAP Wall Area Percentage.

WAR Wall thickness-Artery Ratio.

WHO World Health Organization.

Bibliography

- [1] C. D'Angelis, J. Coalson, and R. Ryan, "Chapter 36 - structure of the respiratory system: Lower respiratory tract," in *Pediatric Critical Care (Fourth Edition)*, B. P. Fuhrman and J. J. Zimmerman, Eds., Fourth Edition, Saint Louis: Mosby, 2011, pp. 490–498. DOI: <https://doi.org/10.1016/B978-0-323-07307-3.10036-9>.
- [2] M. Ochs, J. Nyengaard, A. Jung, L. Knudsen, M. Voigt, T. Wahlers, J. Richter, and H. Gundersen, "The number of alveoli in the human lung," *American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 1, pp. 120–124, 2004. DOI: [10.1164/rccm.200308-11070C](https://doi.org/10.1164/rccm.200308-11070C).
- [3] E. van Rikxoort and B. van Ginneken, "Automated segmentation of pulmonary structures in thoracic computed tomography scans: A review," *Physics in Medicine and Biology*, vol. 58, no. 17, R187–R220, 2013. DOI: [10.1088/0031-9155/58/17/R187](https://doi.org/10.1088/0031-9155/58/17/R187).
- [4] E. Weibel, "Geometry and dimensions of airways of conductive and transitory zones," in *Morphometry of the Human Lung*, Springer, Berlin, Heidelberg, 1963, pp. 110–135. DOI: https://doi.org/10.1007/978-3-642-87553-3_10.
- [5] R. Boaventura, A. Shoemark, and J. Chalmers, "Pathophysiology," in *Bronchiectasis (ERS Monograph)*, J. Chalmers, E. Polverino, and S. Aliberti, Eds., European Respiratory Society, 2018, pp. 8–28. DOI: [10.1183/2312508X.10021218](https://doi.org/10.1183/2312508X.10021218).
- [6] J. Chalmers, A. Chang, S. Chotirmall, R. Dhar, and P. McShane, "Bronchiectasis," *Nature Reviews Disease Primers*, vol. 4, p. 45, 2018. DOI: [10.1038/s41572-018-0042-3](https://doi.org/10.1038/s41572-018-0042-3).
- [7] P. Kumar and M. Clark, "Chapter 24 - respiratory disease," in *Kumar and Clark's Clinical Medicine (Ninth Edition)*, P. Kumar, Ed., Ninth Edition, Edinburgh: Saunders, 2016.
- [8] P. Cole, "Inflammation: A two-edged sword—the model of bronchiectasis," *European journal of respiratory diseases. Supplement*, vol. 147, pp. 6–15, 1986.
- [9] GOLD, "Global initiative for chronic obstructive lung disease (gold). global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease," 2021.
- [10] P. Barnes, P. Burney, E. Silverman, B. Celli, J. Vestbo, J. Wedzicha, and E. Wouters, "Chronic obstructive pulmonary disease," *Nature Reviews Disease Primers*, vol. 1, p. 15 076, 2015. DOI: [10.1038/nrdp.2015.76](https://doi.org/10.1038/nrdp.2015.76).

- [11] *WHO global health estimates - the top 10 causes of death*, "<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>", Accessed: 2022-04-01.
- [12] B. O'Sullivan and S. Freedman, "Cystic fibrosis," *The Lancet*, vol. 373, no. 9678, pp. 1891–1904, 2009. DOI: [https://doi.org/10.1016/S0140-6736\(09\)60327-5](https://doi.org/10.1016/S0140-6736(09)60327-5).
- [13] F. Ratjen, S. Bell, S. Rowe, C. Goss, A. Quittner, and A. Bush, "Cystic fibrosis," *Nature Reviews Disease Primers*, vol. 1, p. 15010, 2015. DOI: 10.1038/nrdp.2015.10.
- [14] H. Kauczor and K. Kreitner, "Mri of the pulmonary parenchyma," *European Radiology*, vol. 9, no. 9, pp. 1755–1764, 1999. DOI: 10.1007/s0033000050919.
- [15] M. Prokop, M. Galanski, A. van der Molen, and C. Schaefer-Prokop, "Spiral and multislice computed tomography of the body," 2003.
- [16] M. Sonka, W. Park, and E. Hoffman, "Rule-based detection of intrathoracic airway trees," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 314–326, 1996. DOI: 10.1109/42.500140.
- [17] W. Kuo, M. de Bruijne, J. Petersen, K. Nasserinejad, H. Ozturk, Y. Chen, A. Perez-Rovira, and H. Tiddens, "Diagnosis of bronchiectasis and airway wall thickening in children with cystic fibrosis: Objective airway-artery quantification," *European Radiology*, vol. 27, no. 11, pp. 4680–4689, 2017. DOI: 10.1007/s00330-017-4819-7.
- [18] J. Pu, S. Gu, S. Liu, S. Zhu, D. Wilson, J. Siegfried, and D. Gur, "CT based computerized identification and analysis of human airways: A review," *Medical Physics*, vol. 39, no. 5, pp. 2603–2616, 2012. DOI: 10.1118/1.4703901.
- [19] B. van Ginneken, "Fifty years of computer analysis in chest imaging: Rule-based, machine learning, deep learning," *Radiological Physics and Technology*, vol. 10, no. 1, pp. 23–32, 2017. DOI: 10.1007/s12194-017-0394-5.
- [20] K. Mori, J. Hasegawa, J. Toriwaki, H. Anno, and K. Katada, "Recognition of bronchus in three-dimensional x-ray ct images with applications to virtualized bronchoscopy system," in *Proceedings 13th International Conference on Pattern Recognition*, vol. 3, 1996, pp. 528–532. DOI: 10.1109/ICPR.1996.547003.
- [21] 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 Transactions on Medical Imaging*, vol. 19, no. 2, pp. 103–114, 2000. DOI: 10.1109/42.836370.
- [22] A. Kiraly, W. Higgins, G. McLennan, E. Hoffman, and J. Reinhardt, "Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy," *Academic Radiology*, vol. 9, no. 10, pp. 1153–1168, 2002. DOI: 10.1016/s1076-6332(03)80517-2.
- [23] D. Aykac, E. Hoffman, G. McLennan, and J. Reinhardt, "Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 940–950, 2003. DOI: 10.1109/TMI.2003.815905.

- [24] C. Fetita, F. Prêteux, C. Beigelman-Aubry, and P. Grenier, “Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 11, pp. 1353–1364, 2004. DOI: 10.1109/TMI.2004.826945.
- [25] A. Fabijańska, “Two-pass region growing algorithm for segmenting airway tree from MDCT chest scans,” *Computational Medical Imaging and Graphics*, vol. 33, no. 7, pp. 537–546, 2009. DOI: 10.1016/j.compmedimag.2009.04.012.
- [26] D. Bartz, D. Mayer, J. Fischer, S. Ley, A. del Río, S. Thust, C. Heussel, H.-U. Kauczor, and W. Strasser, “Hybrid segmentation and exploration of the human lungs,” in *IEEE Visualization, 2003. VIS 2003.*, 2003, pp. 177–184. DOI: 10.1109/VISUAL.2003.1250370.
- [27] D. Mayer, D. Bartz, J. Fischer, S. Ley, A. del Río, S. Thust, H. Kauczor, and C. Heussel, “Hybrid segmentation and virtual bronchoscopy based on CT images,” *Academic Radiology*, vol. 11, no. 5, pp. 551–565, 2004. DOI: 10.1016/j.acra.2004.01.012.
- [28] J. Tschirren, E. Hoffman, G. McLennan, and M. Sonka, “Intrathoracic airway trees: Segmentation and airway morphology analysis from low-dose CT scans,” *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1529–1539, 2005. DOI: 10.1109/TMI.2005.857654.
- [29] B. van Ginneken, W. Baggeman, 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*, 2008, pp. 219–226. DOI: 10.1007/978-3-540-85988-8_27.
- [30] M. Graham, J. Gibbs, D. Cornish, and W. Higgins, “Robust 3-D airway tree segmentation for image-guided peripheral bronchoscopy,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 4, pp. 982–997, 2010. DOI: 10.1109/TMI.2009.2035813.
- [31] C. Bauer, M. Eberlein, and R. Beichel, “Graph-based airway tree reconstruction from chest CT scans: Evaluation of different features on five cohorts,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 5, pp. 1063–1076, 2015. DOI: 10.1109/TMI.2014.2374615.
- [32] 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. DOI: 10.1016/j.media.2010.03.004.
- [33] P. Lo, J. Sparring, J. Pedersen, and M. de Bruijne, “Airway tree extraction with locally optimal paths,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009*, 2009, pp. 51–58. DOI: 10.1007/978-3-642-04271-3_7.
- [34] Z. Bian, J. Charbonnier, J. Liu, D. Zhao, D. Lynch, and B. van Ginneken, “Small airway segmentation in thoracic computed tomography scans: A machine learning approach,” *Physics in Medicine & Biology*, vol. 63, no. 15, p. 155 024, 2018. DOI: 10.1088/1361-6560/aad2a1.

- [35] T. Inoue, Y. Kitamura, Y. Li, and W. Ito, "Robust airway extraction based on machine learning and minimum spanning tree," in *Medical Imaging 2013: Computer-Aided Diagnosis*, International Society for Optics and Photonics, vol. 8670, 2013, pp. 141–149. DOI: 10.1117/12.2006321.
- [36] Q. Meng, T. Kitasaka, Y. Nimura, M. Oda, J. Ueno, and K. Mori, "Automatic segmentation of airway tree based on local intensity filter and machine learning technique in 3D chest CT volume," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 2, pp. 245–261, 2017. DOI: 10.1007/s11548-016-1492-2.
- [37] R. Selvan, J. Petersen, J. Pedersen, and M. de Bruijne, "Extracting tree structures in ct data by tracking multiple statistically ranked hypotheses," *Medical Physics*, vol. 46, no. 10, pp. 4431–4440, 2019. DOI: 10.1002/mp.13711.
- [38] P. Lo, B. van Ginneken, J. Reinhardt, *et al.*, "Extraction of airways from CT (EXACT'09)," *IEEE Transactions on Medical Imaging*, vol. 31, no. 11, pp. 2093–2107, 2012. DOI: 10.1109/TMI.2012.2209674.
- [39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, "<http://www.deeplearningbook.org>".
- [40] G. Litjens, T. Kooi, B. Bejnordi, A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, and C. Sanchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017. DOI: <https://doi.org/10.1016/j.media.2017.07.005>.
- [41] A. Garcia-Uceda, R. Selvan, Z. Saghir, H. Tiddens, and M. de Bruijne, "Automatic airway segmentation from computed tomography using robust and efficient 3-D convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, p. 16001, 2021. DOI: 10.1038/s41598-021-95364-1.
- [42] A. Garcia-Uceda Juarez, R. Selvan, Z. Saghir, and M. de Bruijne, "A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs," in *Machine Learning in Medical Imaging*, 2019, pp. 583–591. DOI: 10.1007/978-3-030-32692-0_67.
- [43] J. Charbonnier, E. van Rikxoort, A. Setio, C. Schaefer-Prokop, B. van Ginneken, and F. Ciompi, "Improving airway segmentation in computed tomography using leak detection with convolutional networks," *Medical Image Analysis*, vol. 36, pp. 52–60, 2017. DOI: <https://doi.org/10.1016/j.media.2016.11.001>.
- [44] J. Yun, J. Park, D. Yu, J. Yi, M. Lee, H. Park, J. Lee, J. Seo, and N. Kim, "Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net," *Medical Image Analysis*, vol. 51, pp. 13–20, 2019. DOI: <https://doi.org/10.1016/j.media.2018.10.006>.
- [45] D. Jin, Z. Xu, A. Harrison, K. George, and D. Mollura, "3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels," in *Machine Learning in Medical Imaging*, 2017, pp. 141–149. DOI: 10.1007/978-3-319-67389-9_17.

- [46] Q. Meng, H. Roth, T. Kitasaka, M. Oda, J. Ueno, and K. Mori, "Tracking and segmentation of the airways in chest CT using a fully convolutional network," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, 2017, pp. 198–207. DOI: 10.1007/978-3-319-66185-8_23.
- [47] Y. Qin, M. Chen, H. Zheng, Y. Gu, M. Shen, J. Yang, X. Huang, Y. Zhu, and G. Yang, "AirwayNet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 212–220. DOI: 10.1007/978-3-030-32226-7_24.
- [48] Y. Qin, Y. Gu, H. Zheng, M. Chen, J. Yang, and Y. Zhu, "AirwayNet-SE: A simple-yet-effective approach to improve airway segmentation using context scale fusion," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 809–813. DOI: 10.1109/ISBI45749.2020.9098537.
- [49] T. Zhao, Z. Yin, J. Wang, D. Gao, Y. Chen, and Y. Mao, "Bronchus segmentation and classification by neural networks and linear programming," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 230–239. DOI: 10.1007/978-3-030-32226-7_26.
- [50] C. Wang, Y. Hayashi, M. Oda, H. Itoh, T. Kitasaka, A. Frangi, and K. Mori, "Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 348–356. DOI: 10.1007/978-3-030-32226-7_39.
- [51] Y. Qin, H. Zheng, Y. Gu, X. Huang, J. Yang, L. Wang, and Y. Zhu, "Learning bronchiole-sensitive airway segmentation CNNs by feature recalibration and attention distillation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2020*, 2020, pp. 221–231. DOI: 10.1007/978-3-030-59710-8_22.
- [52] Y. Qin, H. Zheng, Y. Gu, X. Huang, J. Yang, L. Wang, F. Yao, Y. Zhu, and G. Yang, "Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT," *IEEE Transactions on Medical Imaging*, vol. 40, no. 6, pp. 1603–1617, 2021. DOI: 10.1109/TMI.2021.3062280.
- [53] H. Zheng, Y. Qin, Y. Gu, F. Xie, J. Yang, J. Sun, and G. Yang, "Alleviating class-wise gradient imbalance for pulmonary airway segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2452–2462, 2021. DOI: 10.1109/TMI.2021.3078828.
- [54] K. Zhou, N. Chen, X. Xu, Z. Wang, J. Guo, L. Liu, and Z. Yi, "Automatic airway tree segmentation based on multi-scale context information," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 2, pp. 219–230, 2021. DOI: 10.1007/s11548-020-02293-x.
- [55] S. Nadeem, E. Hoffman, J. Sieren, A. Comellas, S. Bhatt, I. Barjaktarevic, F. Abtin, and P. Saha, "A CT-based automated algorithm for airway segmentation using freeze-and-grow propagation and deep learning," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 405–418, 2021. DOI: 10.1109/TMI.2020.3029013.

- [56] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, pp. 234–241, 2015. DOI: 10.1007/978-3-319-24574-4_28.
- [57] X. Liu, D. Chen, M. Tawhai, X. Wu, E. Hoffman, and M. Sonka, “Optimal graph search based segmentation of airway tree double surfaces across bifurcations,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 3, pp. 493–510, 2013. DOI: 10.1109/TMI.2012.2223760.
- [58] J. Petersen, M. Nielsen, P. Lo, L. Nordenmark, J. Pedersen, M. 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*, vol. 18, no. 3, pp. 531–541, 2014. DOI: <https://doi.org/10.1016/j.media.2014.02.004>.
- [59] F. Mohamed Hoesein, P. de Jong, J. Lammers, *et al.*, “Airway wall thickness associated with forced expiratory volume in 1 second decline and development of airflow limitation,” *European Respiratory Journal*, vol. 45, no. 3, pp. 644–651, 2015. DOI: 10.1183/09031936.00020714.
- [60] P. Konietzke, O. Weinheimer, W. Wagner, F. Wuennemann, C. Hintze, J. Biederer, C. Heussel, H. Kauczor, and M. Wielpütz, “Optimizing airway wall segmentation and quantification by reducing the influence of adjacent vessels and intravascular contrast material with a modified integral-based algorithm in quantitative computed tomography,” *PLOS ONE*, vol. 15, no. 8, p. e0237939, 2020. DOI: 10.1371/journal.pone.0237939.
- [61] J. Petersen, M. Wille, L. Rakêt, 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. DOI: 10.1007/s00330-014-3261-3.
- [62] A. Perez-Rovira, W. Kuo, J. Petersen, H. Tiddens, and M. de Bruijne, “Automatic airway-artery analysis on lung CT to quantify airway wall thickening and bronchiectasis,” *Medical Physics*, vol. 43, no. 10, pp. 5736–5744, 2016. DOI: 10.1118/1.4963214.
- [63] T. Schlathoelter, C. Lorenz, I. Carlsen, S. Renisch, and T. Deschamps, “Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy,” in *Medical Imaging 2002: Image Processing*, vol. 4684, SPIE, 2002, pp. 103–113. DOI: 10.1117/12.467061.
- [64] Y. Nakano, J. Wong, P. de Jong, L. Buzatu, T. Nagao, H. Coxson, W. Elliott, J. Hogg, and P. Pare, “The prediction of small airway dimensions using computed tomography,” *American journal of respiratory and critical care medicine*, vol. 171, no. 2, pp. 142–146, 2005. DOI: 10.1164/rccm.200407-8740C.
- [65] W. Kuo, A. Perez-Rovira, H. Tiddens, *et al.*, “Airway tapering: An objective image biomarker for bronchiectasis,” *European Radiology*, vol. 30, no. 5, pp. 2703–2711, 2020. DOI: 10.1007/s00330-019-06606-w.

- [66] K. Quan, J. Jacob, R. Shipley, D. Hawkes, and J. Hurst, "Airway tapering in bronchiectatic and healthy airways," *European Respiratory Journal*, vol. 52, OA3793, 2018. DOI: 10.1183/13993003.congress-2018.OA3793.
- [67] K. Quan, R. Tanno, R. Shipley, J. Brown, J. Jacob, J. Hurst, and D. Hawkes, "Reproducibility of an airway tapering measurement in computed tomography with application to bronchiectasis," *Journal of Medical Imaging*, vol. 6, no. 3, p. 034003, 2019. DOI: 10.1117/1.JMI.6.3.034003.
- [68] W. Kuo, T. Soffers, E. Andrinopoulou, T. Rosenow, S. Ranganathan, L. Turkovic, S. Stick, H. Tiddens, and on behalf of AREST CF, "Quantitative assessment of airway dimensions in young children with cystic fibrosis lung disease using chest computed tomography," *Pediatric Pulmonology*, vol. 52, no. 11, pp. 1414–1423, 2017. DOI: <https://doi.org/10.1002/ppul.23787>.
- [69] W. Kuo, E. Andrinopoulou, A. Perez-Rovira, H. Ozturk, M. de Bruijne, and H. Tiddens, "Objective airway artery dimensions compared to CT scoring methods assessing structural cystic fibrosis lung disease," *Journal of Cystic Fibrosis*, vol. 16, no. 1, pp. 116–123, 2017. DOI: <https://doi.org/10.1016/j.jcf.2016.05.015>.
- [70] Z. Naseri, S. Sherafat, H. Abrishami Moghaddam, M. Modaresi, N. Pak, and F. Zamani, "Semi-automatic methods for airway and adjacent vessel measurement in bronchiectasis patterns in lung HRCT images of cystic fibrosis patients," *Journal of Digital Imaging*, vol. 31, no. 5, pp. 727–737, 2018. DOI: 10.1007/s10278-018-0076-9.
- [71] Q. Lv, R. Sandvik, K. Nielsen, E. Andrinopoulou, L. Gallardo-Estrella, J. Charbonnier, and H. Tiddens, "Sensitive automated airway-artery method to monitor progression of CF airway disease," *European Respiratory Journal*, vol. 58, OA2676, 2021. DOI: 10.1183/13993003.congress-2021.OA2676.
- [72] A. Nambu, J. Zach, J. Schroeder, G. Jin, S. Kim, Y. Kim, C. Schnell, R. Bowler, and D. Lynch, "Quantitative computed tomography measurements to evaluate airway disease in chronic obstructive pulmonary disease: Relationship to physiological measurements, clinical index and visual assessment of airway disease," *European Journal of Radiology*, vol. 85, no. 11, pp. 2144–2151, 2016. DOI: <https://doi.org/10.1016/j.ejrad.2016.09.010>.
- [73] J. Charbonnier, E. Pompe, C. Moore, S. Humphries, B. van Ginneken, B. Make, E. Regan, J. Crapo, E. van Rikxoort, and D. Lynch, "Airway wall thickening on CT: Relation to smoking status and severity of COPD," *Respiratory Medicine*, vol. 146, pp. 36–41, 2019. DOI: <https://doi.org/10.1016/j.rmed.2018.11.014>.
- [74] G. McGuinness, D. Naidich, B. Leitman, and D. McCauley, "Bronchiectasis: CT evaluation," *American Journal of Roentgenology*, vol. 160, no. 2, pp. 253–259, 1993. DOI: 10.2214/ajr.160.2.8424327.
- [75] J. Tschirren, T. Yavarna, and J. Reinhardt, "Airway segmentation framework for clinical environments," in *The Second International Workshop on Pulmonary Image Analysis*, 2009, pp. 227–238.

- [76] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.
- [77] A. Garcia-Uceda Juarez, H. Tiddens, and M. de Bruijne, "Automatic airway segmentation in chest CT using convolutional neural networks," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, 2018, pp. 238–250. DOI: 10.1007/978-3-030-00946-5_24.
- [78] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2020*, 2016, pp. 424–432. DOI: 10.1007/978-3-319-46723-8_49.
- [79] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565–571. DOI: 10.1109/3DV.2016.79.
- [80] R. Selvan, T. Kipf, M. Welling, A. Garcia-Uceda Juarez, J. Pedersen, J. Petersen, and M. de Bruijne, "Graph refinement based airway extraction using mean-field networks and graph neural networks," *Medical Image Analysis*, vol. 64, p. 101751, 2020. DOI: <https://doi.org/10.1016/j.media.2020.101751>.
- [81] C. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2017, pp. 240–248. DOI: 10.1007/978-3-319-67558-9_28.
- [82] A. Paszke, S. Gross, F. Massa, *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019, pp. 8024–8035.
- [83] C. Fiorio and J. Gustedt, "Two linear time union-find strategies for image processing," *Theoretical Computer Science*, vol. 154, no. 2, pp. 165–181, 1996. DOI: [https://doi.org/10.1016/0304-3975\(94\)00262-2](https://doi.org/10.1016/0304-3975(94)00262-2).
- [84] J. Pedersen, H. Ashraf, A. Dirksen, *et al.*, "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, 2009. DOI: <https://doi.org/10.1097/JTO.0b013e3181a0d98f>.
- [85] *Extraction of airways from CT 2009 (EXACT'09)*, "http://image.diku.dk/exact/new_results.php".
- [86] D. Gil, C. Sanchez, A. Borrás, M. Diez-Ferrer., and A. Rosell, "Segmentation of distal airways using structural analysis," *PLOS ONE*, vol. 14, no. 12, p. e0226006, 2019. DOI: 10.1371/journal.pone.0226006.
- [87] F. Isensee, P. Jaeger, S. Kohl, J. Petersen, and K. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021. DOI: 10.1038/s41592-020-01008-z.

- [88] D. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [89] T. Lee, R. Kashyap, and C. Chu, “Building skeleton models via 3-D medial surface axis thinning algorithms,” *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994. DOI: <https://doi.org/10.1006/cgip.1994.1042>.
- [90] E. van Rikxoort, W. Baggeman, and B. van Ginneken, “Automatic segmentation of the airway tree from thoracic CT scans using a multithreshold approach,” in *The Second International Workshop on Pulmonary Image Analysis*, 2009, pp. 341–349.
- [91] R. Estépar, J. Ross, K. Russian, T. Schultz, G. Washko, and G. Kindlmann, “Computational vascular morphometry for the assessment of pulmonary vascular disease based on scale-space particles,” in *2012 IEEE 9th International Symposium on Biomedical Imaging (ISBI)*, 2012, pp. 1479–1482. DOI: [10.1109/ISBI.2012.6235851](https://doi.org/10.1109/ISBI.2012.6235851).
- [92] R. Brügger, C. Baumgartner, and E. Konukoglu, “A partially reversible U-Net for memory-efficient volumetric image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 429–437. DOI: [10.1007/978-3-030-32248-9_48](https://doi.org/10.1007/978-3-030-32248-9_48).
- [93] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, “Densely connected convolutional networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269. DOI: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [94] N. Siddique, S. Paheding, C. Elkin, and V. Devabhaktuni, “U-Net and its variants for medical image segmentation: A review of theory and applications,” *IEEE Access*, vol. 9, pp. 82 031–82 057, 2021. DOI: [10.1109/ACCESS.2021.3086020](https://doi.org/10.1109/ACCESS.2021.3086020).
- [95] J. Su, L. Wolff, A. van Es, W. van Zwam, C. Majoie, D. Dippel, A. van der Lugt, W. Niessen, and T. van Walsum, “Automatic collateral scoring from 3D CTA images,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2190–2200, 2020. DOI: [10.1109/TMI.2020.2966921](https://doi.org/10.1109/TMI.2020.2966921).
- [96] D. Jha, M. Riegler, D. Johansen, P. Halvorsen, and H. Johansen, “DoubleU-Net: A deep convolutional neural network for medical image segmentation,” in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020, pp. 558–564. DOI: [10.1109/CBMS49503.2020.00111](https://doi.org/10.1109/CBMS49503.2020.00111).
- [97] W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, and E. Xing, “SCAN: Structure correcting adversarial network for organ segmentation in chest x-rays,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 263–273. DOI: [10.1007/978-3-030-00889-5_30](https://doi.org/10.1007/978-3-030-00889-5_30).
- [98] Y. Yang, Z. Wang, J. Liu, K. Cheng, and X. Yang, *Label refinement with an iterative generative adversarial network for boosting retinal vessel segmentation*, 2019. arXiv: [1912.02589](https://arxiv.org/abs/1912.02589).

- [99] R. Araújo, J. Cardoso, and H. Oliveira, "A deep learning design for improving topology coherence in blood vessel segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 93–101. DOI: 10.1007/978-3-030-32239-7_11.
- [100] G. Cheng, X. Wu, W. Xiang, C. Guo, H. Ji, and L. He, "Segmentation of the airway tree from chest CT using tiny atrous convolutional network," *IEEE Access*, vol. 9, pp. 33 583–33 594, 2021. DOI: 10.1109/ACCESS.2021.3059680.
- [101] H. Zheng, Y. Qin, Y. Gu, F. Xie, J. Sun, J. Yang, and G. Yang, "Refined local-imbalance-based weight for airway segmentation in CT," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2021*, 2021, pp. 410–419. DOI: 10.1007/978-3-030-87193-2_39.
- [102] P. Sanches, C. Meyer, V. Vigon, and B. Naegel, "Cerebrovascular network segmentation of MRA images with deep learning," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 768–771. DOI: 10.1109/ISBI.2019.8759569.
- [103] M. Livne, J. Rieger, O. Aydin, *et al.*, "A U-Net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease," *Frontiers in Neuroscience*, vol. 13, 2019. DOI: 10.3389/fnins.2019.00097.
- [104] A. Hilbert, V. Madai, E. Akay, *et al.*, "BRAVE-NET: Fully automated arterial brain vessel segmentation in patients with cerebrovascular disease," *Frontiers in Artificial Intelligence*, vol. 3, 2020. DOI: 10.3389/frai.2020.552258.
- [105] M. Meijjs, A. Patel, S. van de Leemput, M. Prokop, E. van Dijk, F. de Leeuw, F. Meijer, B. van Ginneken, and R. Manniesing, "Robust segmentation of the full cerebral vasculature in 4D CT of suspected stroke patients," *Scientific Reports*, vol. 7, no. 1, p. 15 622, 2017. DOI: 10.1038/s41598-017-15617-w.
- [106] N. Bouma, H. Janssens, E. Andrinopoulou, and H. Tiddens, "Airway disease on chest computed tomography of preschool children with cystic fibrosis is associated with school-age bronchiectasis," *Pediatric Pulmonology*, vol. 55, no. 1, pp. 141–148, 2020. DOI: <https://doi.org/10.1002/ppul.24498>.
- [107] I. Jansen, M. Mulder, and R. Goldhoorn, "Endovascular treatment for acute ischaemic stroke in routine clinical practice: Prospective, observational cohort study (MR CLEAN registry)," *BMJ*, vol. 360, 2018. DOI: 10.1136/bmj.k949.
- [108] D. Rodriguez-Luna, D. Dowlathshahi, R. Aviv, *et al.*, "Venous phase of computed tomography angiography increases spot sign detection, but intracerebral hemorrhage expansion is greater in spot signs detected in arterial phase," *Stroke*, vol. 45, no. 3, pp. 734–739, 2014. DOI: 10.1161/STROKEAHA.113.003007.
- [109] R. Peter, B. Emmer, A. van Es, and T. van Walsum, "Cortical and vascular probability maps for analysis of human brain in computed tomography images," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 1141–1145. DOI: 10.1109/ISBI.2017.7950718.

- [110] H. Tiddens, S. Donaldson, M. Rosenfeld, and P. Pare, “Cystic fibrosis lung disease starts in the small airways: Can we treat it more effectively?” *Pediatric Pulmonology*, vol. 45, no. 2, pp. 107–117, 2010. DOI: <https://doi.org/10.1002/ppul.21154>.
- [111] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [112] G. Bortsova, F. Dubost, L. Hogeweg, I. Katramados, and M. de Bruijne, “Semi-supervised medical image segmentation via learning consistency under transformations,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 810–818. DOI: 10.1007/978-3-030-32226-7_90.
- [113] S. Chen, Z. Sedghi Gamechi, F. Dubost, G. van Tulder, and M. de Bruijne, “An end-to-end approach to segmentation in medical images with CNN and posterior-CRF,” *Medical Image Analysis*, vol. 76, p. 102311, 2022. DOI: <https://doi.org/10.1016/j.media.2021.102311>.
- [114] F. Scarselli, M. Gori, A. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009. DOI: 10.1109/TNN.2008.2005605.
- [115] T. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *International Conference on Learning Representations*, 2017.
- [116] R. Selvan, T. Kipf, M. Welling, J. Pedersen, J. Petersen, and M. de Bruijne, “Extraction of airways using graph neural networks,” in *1st Conference on Medical Imaging with Deep Learning - MIDL 2018*, 2018.
- [117] S. Shin, S. Lee, I. Yun, and K. Lee, “Deep vessel segmentation by learning graphical connectivity,” *Medical Image Analysis*, vol. 58, p. 101556, 2019. DOI: <https://doi.org/10.1016/j.media.2019.101556>.
- [118] K. Lowe, E. Regan, A. Anzueto, *et al.*, “COPDGene 2019: Redefining the diagnosis of chronic obstructive pulmonary disease,” *Chronic Obstructive Pulmonary Diseases*, vol. 6, no. 5, pp. 384–399, 2019. DOI: 10.15326/jcopdf.6.5.2019.0149.
- [119] E. Hammond, C. Sloan, J. Newell Jr., *et al.*, “Comparison of low- and ultralow-dose computed tomography protocols for quantitative lung and airway assessment,” *Medical Physics*, vol. 44, no. 9, pp. 4747–4757, 2017. DOI: <https://doi.org/10.1002/mp.12436>.
- [120] S. Quaderi and J. Hurst, “The unmet global burden of COPD,” *Global Health, Epidemiology and Genomics*, vol. 3, p. e4, 2018. DOI: 10.1017/ghg.2018.1.
- [121] C. Xia, M. Rook, G. Pelgrim, *et al.*, “Early imaging biomarkers of lung cancer, COPD and coronary artery disease in the general population: Rationale and design of the ImaLife (imaging in lifelines) study,” *European Journal of Epidemiology*, vol. 35, pp. 75–86, 2020. DOI: 10.1007/s10654-019-00519-0.
- [122] D. Xu, H. Gietema, H. de Koning, *et al.*, “Nodule management protocol of the NELSON randomised lung cancer screening trial,” *Lung Cancer*, vol. 54, no. 2, pp. 177–184, 2006. DOI: <https://doi.org/10.1016/j.lungcan.2006.08.006>.

- [123] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, *et al.*, “3D Slicer as an image computing platform for the quantitative imaging network,” *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012. DOI: <https://doi.org/10.1016/j.mri.2012.05.001>.
- [124] P. Virtanen, R. Gommers, T. Oliphant, *et al.*, “SciPy 1.0: Fundamental algorithms for scientific computing in python,” *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [125] W. van den Bosch, A. James, and H. Tiddens, “Structure and function of small airways in asthma patients revisited,” *European Respiratory Review*, vol. 30, no. 159, p. 200186, 2021. DOI: [10.1183/16000617.0186-2020](https://doi.org/10.1183/16000617.0186-2020).
- [126] K. Gove, T. Wilkinson, S. Jack, K. Ostridge, B. Thompson, and J. Conway, “Systematic review of evidence for relationships between physiological and CT indices of small airways and clinical outcomes in COPD,” *Respiratory Medicine*, vol. 139, pp. 117–125, 2018. DOI: <https://doi.org/10.1016/j.rmed.2018.05.005>.
- [127] R. Eddy, S. Svenningsen, M. Kirby, D. Knipping, D. McCormack, C. Licskai, P. Nair, and G. Parraga, “Is computed tomography airway count related to asthma severity and airway structure and function?” *American journal of respiratory and critical care medicine*, vol. 201, no. 8, pp. 923–933, 2020. DOI: [10.1164/rccm.201908-15520C](https://doi.org/10.1164/rccm.201908-15520C).
- [128] T. Oguma, T. Hirai, M. Fukui, *et al.*, “Longitudinal shape irregularity of airway lumen assessed by CT in patients with bronchial asthma and COPD,” *Thorax*, vol. 70, no. 8, pp. 719–724, 2015. DOI: [10.1136/thoraxjnl-2014-206651](https://doi.org/10.1136/thoraxjnl-2014-206651).
- [129] L. Prevedello, S. Halabi, G. Shih, C. Wu, M. Kohli, F. Chokshi, B. Erickson, J. Kalpathy-Cramer, K. Andriole, and A. Flanders, “Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions,” *Radiology: Artificial Intelligence*, vol. 1, no. 1, p. e180031, 2019. DOI: [10.1148/ryai.2019180031](https://doi.org/10.1148/ryai.2019180031).
- [130] M. Heuvelmans, M. Vonder, M. Rook, H. Groen, G. De Bock, X. Xie, M. Ijzerman, R. Vliegenthart, and M. Oudkerk, “Screening for early lung cancer, chronic obstructive pulmonary disease, and cardiovascular disease (the big-3) using low-dose chest computed tomography: Current evidence and technical considerations,” *Journal of Thoracic Imaging*, vol. 34, no. 3, pp. 160–169, 2019. DOI: [10.1097/RTI.0000000000000379](https://doi.org/10.1097/RTI.0000000000000379).
- [131] J. Reinhardt, N. D’Souza, and E. Hoffman, “Accurate measurement of intrathoracic airways,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 820–827, 1997. DOI: [10.1109/42.650878](https://doi.org/10.1109/42.650878).
- [132] R. Estépar, G. Washko, E. Silverman, J. Reilly, R. Kikinis, and C. Westin, “Accurate airway wall estimation using phase congruency,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006*, 2006, pp. 125–134. DOI: [10.1007/11866763_16](https://doi.org/10.1007/11866763_16).

- [133] I. Dudurych, A. Garcia-Uceda, Z. Saghir, H. Tiddens, R. Vliegenthart, and M. de Bruijne, "Creating a training set for artificial intelligence from initial segmentations of airways," *European Radiology Experimental*, vol. 5, no. 1, p. 54, 2021. DOI: 10.1186/s41747-021-00247-9.
- [134] J. Sieren, K. Gunderson, D. Lynch, J. Newell, P. Judy, and E. Hoffman, "COPDGene phantom: Quality control of quantitative lung imaging in a multicenter trial," in *D27. PUSHING IMAGING TOWARD MICROANATOMY AND ORGAN/CELL PHYSIOLOGY*. American Thoracic Society, 2010, A5519–A5519. DOI: 10.1164/ajrccm-conference.2010.181.1_MeetingAbstracts.A5519.
- [135] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631. DOI: 10.1145/3292500.3330701.
- [136] D. Merkel, "Docker: Lightweight linux containers for consistent development and deployment," *Linux J.*, vol. 2014, no. 239, 2014. DOI: 10.5555/2600239.2600241.
- [137] A. Sijtsma, J. Rienks, P. van der Harst, G. Navis, J. Rosmalen, and A. Dotinga, "Cohort profile update: Lifelines, a three-generation cohort study and biobank," *International Journal of Epidemiology*, dyab257, 2021. DOI: <https://doi.org/10.1093/ije/dyab257>.
- [138] K. Kuwano, C. Bosken, P. Paré, T. Bai, B. Wiggs, and J. Hogg, "Small airways dimensions in asthma and in chronic obstructive pulmonary disease," *American Review of Respiratory Disease*, vol. 148, no. 5, pp. 1220–1225, 1993. DOI: 10.1164/ajrccm/148.5.1220.
- [139] I. Dudurych, S. Muiser, N. McVeigh, H. Kerstjens, M. van den Berge, M. de Bruijne, and R. Vliegenthart, "Bronchial wall parameters on CT in healthy never-smoking, smoking, COPD, and asthma populations: A systematic review and meta-analysis," *European Radiology*, 2022. DOI: 10.1007/s00330-022-08600-1.
- [140] Z. Xu, U. Bagci, B. Foster, A. Mansoor, J. Udupa, and D. Mollura, "A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT," *Medical Image Analysis*, vol. 24, no. 1, pp. 1–17, 2015. DOI: <https://doi.org/10.1016/j.media.2015.05.003>.
- [141] G. King, N. Müller, K. Whittall, Q. Xiang, and P. Paré, "An analysis algorithm for measuring airway lumen and wall areas from high-resolution computed tomographic data," *American Journal of Respiratory and Critical Care Medicine*, vol. 161, no. 2, pp. 574–580, 2000. DOI: 10.1164/ajrccm.161.2.9812073.
- [142] P. Davis, "Cystic fibrosis since 1938," *American Journal of Respiratory and Critical Care Medicine*, vol. 173, no. 5, pp. 475–482, 2006. DOI: 10.1164/rccm.200505-8400E.

- [143] L. Mott, J. Park, C. Murray, *et al.*, “Progression of early structural lung disease in young children with cystic fibrosis assessed using CT,” *Thorax*, vol. 67, no. 6, pp. 509–516, 2012. DOI: 10.1136/thoraxjnl-2011-200912.
- [144] S. Stick, S. Brennan, C. Murray, *et al.*, “Bronchiectasis in infants and preschool children diagnosed with cystic fibrosis after newborn screening,” *The Journal of Pediatrics*, vol. 155, no. 5, pp. 623–628, 2009. DOI: 10.1016/j.jpeds.2009.05.005.
- [145] P. Sly, C. Gangell, L. Chen, R. Ware, S. Ranganathan, L. Mott, C. Murray, and S. Stick, “Risk factors for bronchiectasis in children with cystic fibrosis,” *New England Journal of Medicine*, vol. 368, no. 21, pp. 1963–1970, 2013. DOI: 10.1056/NEJMoa1301725.
- [146] H. Tiddens, E. Andrinopoulou, J. McIntosh, J. Elborn, E. Kerem, N. Bouma, J. Bosch, and M. Kemner-van de Corput, “Chest computed tomography outcomes in a randomized clinical trial in cystic fibrosis: Lessons learned from the first ataluren phase 3 study,” *PLOS ONE*, vol. 15, no. 11, p. e0240898, 2020. DOI: 10.1371/journal.pone.0240898.
- [147] B. Ramsey, J. Davies, N. McElvaney, *et al.*, “A CFTR potentiator in patients with cystic fibrosis and the G551D mutation,” *New England Journal of Medicine*, vol. 365, no. 18, pp. 1663–1672, 2011. DOI: 10.1056/NEJMoa1105185.
- [148] M. Boyle, S. Bell, M. Konstan, S. McColley, S. Rowe, E. Rietschel, X. Huang, D. Waltz, N. Patel, and D. Rodman, “A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: A phase 2 randomised controlled trial,” *The Lancet Respiratory Medicine*, vol. 2, no. 7, pp. 527–538, 2014. DOI: 10.1016/S2213-2600(14)70132-8.
- [149] M. Rosenfeld, C. Wainwright, M. Higgins, *et al.*, “Ivacaftor treatment of cystic fibrosis in children aged 12 to <24 months and with a CFTR gating mutation (ARRIVAL): A phase 3 single-arm study,” *The Lancet Respiratory Medicine*, vol. 6, no. 7, pp. 545–553, 2018. DOI: 10.1016/S2213-2600(18)30202-9.
- [150] P. Flume, C. Wainwright, D. Elizabeth Tullis, S. Rodriguez, M. Niknian, M. Higgins, J. Davies, and J. Wagener, “Recovery of lung function following a pulmonary exacerbation in patients with cystic fibrosis and the G551D-CFTR mutation treated with ivacaftor,” *Journal of Cystic Fibrosis*, vol. 17, no. 1, pp. 83–88, 2018. DOI: 10.1016/j.jcf.2017.06.002.
- [151] S. McColley, M. Konstan, B. Ramsey, *et al.*, “Lumacaftor/ivacaftor reduces pulmonary exacerbations in patients irrespective of initial changes in fev1,” *Journal of Cystic Fibrosis*, vol. 18, no. 1, pp. 94–101, 2019. DOI: 10.1016/j.jcf.2018.07.011.
- [152] K. Hisert, S. Heltshe, C. Pope, *et al.*, “Restoring cystic fibrosis transmembrane conductance regulator function reduces airway bacteria and inflammation in people with cystic fibrosis and chronic lung infections,” *American Journal of Respiratory and Critical Care Medicine*, vol. 195, no. 12, pp. 1617–1628, 2017. DOI: 10.1164/rccm.201609-19540C.

- [153] D. Hubert, C. Dehillotte, A. Munck, *et al.*, “Retrospective observational study of french patients with cystic fibrosis and a Gly551Asp-CFTR mutation after 1 and 2 years of treatment with ivacaftor in a real-world setting,” *Journal of Cystic Fibrosis*, vol. 17, no. 1, pp. 89–95, 2018. DOI: <https://doi.org/10.1016/j.jcf.2017.07.001>.
- [154] F. Ratjen, C. Hug, G. Marigowda, *et al.*, “Efficacy and safety of lumacaftor and ivacaftor in patients aged 6-11 years with cystic fibrosis homozygous for F508del-CFTR: A randomised, placebo-controlled phase 3 trial,” *The Lancet Respiratory Medicine*, vol. 5, no. 7, pp. 557–567, 2017. DOI: [https://doi.org/10.1016/S2213-2600\(17\)30215-1](https://doi.org/10.1016/S2213-2600(17)30215-1).
- [155] S. Donaldson, B. Laube, T. Corcoran, *et al.*, “Effect of ivacaftor on mucociliary clearance and clinical outcomes in cystic fibrosis patients with G551D-CFTR,” *JCI Insight*, vol. 3, no. 24, p. e122695, 2018. DOI: [10.1172/jci.insight.122695](https://doi.org/10.1172/jci.insight.122695).
- [156] K. De Boeck, A. Munck, S. Walker, A. Faro, P. Hiatt, G. Gilmartin, and M. Higgins, “Efficacy and safety of ivacaftor in patients with cystic fibrosis and a non-G551D gating mutation,” *Journal of Cystic Fibrosis*, vol. 13, no. 6, pp. 674–680, 2014. DOI: <https://doi.org/10.1016/j.jcf.2014.09.005>.
- [157] C. Wainwright, J. Elborn, B. Ramsey, *et al.*, “Lumacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del CFTR,” *New England Journal of Medicine*, vol. 373, no. 3, pp. 220–231, 2015. DOI: [10.1056/NEJMoa1409547](https://doi.org/10.1056/NEJMoa1409547).
- [158] S. Sheikh, F. Long, K. McCoy, T. Johnson, N. Ryan-Wenger, and D. Hayes, “Computed tomography correlates with improvement with ivacaftor in cystic fibrosis patients with G551D mutation,” *Journal of Cystic Fibrosis*, vol. 14, no. 1, pp. 84–89, 2015. DOI: <https://doi.org/10.1016/j.jcf.2014.06.011>.
- [159] G. Chassagnon, D. Hubert, I. Fajac, P. Burgel, and M. Revel, “Long-term computed tomographic changes in cystic fibrosis patients treated with ivacaftor,” *European Respiratory Journal*, vol. 48, no. 1, pp. 249–252, 2016. DOI: [10.1183/13993003.01918-2015](https://doi.org/10.1183/13993003.01918-2015).
- [160] N. Ronan, G. Einarsson, M. Twomey, *et al.*, “CORK study in cystic fibrosis: Sustained improvements in ultra-low-dose chest CT scores after CFTR modulation with ivacaftor,” *Chest*, vol. 153, no. 2, pp. 395–403, 2018. DOI: <https://doi.org/10.1016/j.chest.2017.10.005>.
- [161] A. Brody, S. Nagle, C. Hug, G. Marigowda, D. Waltz, J. Goldin, F. Ratjen, and L. Wang, “S93 effect of lumacaftor/ivacaftor on total, bronchiectasis, and air trapping computed tomography (CT) scores in children homozygous for f508del-CFTR: Exploratory imaging substudy,” *Thorax*, vol. 72, no. Suppl 3, A57–A57, 2017. DOI: [10.1136/thoraxjnl-2017-210983.99](https://doi.org/10.1136/thoraxjnl-2017-210983.99).
- [162] G. Dournes, C. Hall, M. Willmering, *et al.*, “Artificial intelligence in CT for quantifying lung changes in the era of CFTR modulators,” *European Respiratory Journal*, p. 2100844, 2021. DOI: [10.1183/13993003.00844-2021](https://doi.org/10.1183/13993003.00844-2021).

- [163] S. Ranganathan, C. Dezateux, A. Bush, *et al.*, “Airway function in infants newly diagnosed with cystic fibrosis,” *The Lancet*, vol. 358, no. 9297, pp. 1964–1965, 2001. DOI: [https://doi.org/10.1016/S0140-6736\(01\)06970-7](https://doi.org/10.1016/S0140-6736(01)06970-7).
- [164] K. Ramsey, T. Rosenow, L. Turkovic, *et al.*, “Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis,” *American Journal of Respiratory and Critical Care Medicine*, vol. 193, no. 1, pp. 60–67, 2016. DOI: [10.1164/rccm.201507-14090C](https://doi.org/10.1164/rccm.201507-14090C).
- [165] P. de Jong, A. Lindblad, L. Rubin, W. Hop, J. de Jongste, M. Brink, and H. Tiddens, “Progression of lung disease on computed tomography and pulmonary function tests in children and adults with cystic fibrosis,” *Thorax*, vol. 61, no. 1, pp. 80–85, 2006. DOI: [10.1136/thx.2005.045146](https://doi.org/10.1136/thx.2005.045146).
- [166] T. Rosenow, M. Oudraad, C. Murray, L. Turkovic, W. Kuo, M. de Bruijne, S. Ranganathan, H. Tiddens, and S. Stick, “PRAGMA-CF. a quantitative structural lung disease computed tomography outcome in young children with cystic fibrosis,” *American Journal of Respiratory and Critical Care Medicine*, vol. 191, no. 10, pp. 1158–1165, 2015. DOI: [10.1164/rccm.201501-00610C](https://doi.org/10.1164/rccm.201501-00610C).
- [167] J. Otjen, J. Swanson, A. Oron, R. DiBlasi, T. Swortzel, J. van Well, E. Gommers, and M. Rosenfeld, “Spirometry-assisted high resolution chest computed tomography in children: Is it worth the effort?” *Current Problems in Diagnostic Radiology*, vol. 47, no. 1, pp. 14–18, 2018. DOI: <https://doi.org/10.1067/j.cpradiol.2017.02.010>.
- [168] A. Brody, M. Kosorok, Z. Li, L. Broderick, J. Foster, A. Laxova, H. Bandla, and P. Farrell, “Reproducibility of a scoring system for computed tomography scanning in cystic fibrosis,” *Journal of Thoracic Imaging*, vol. 21, no. 1, pp. 14–21, 2006. DOI: [10.1097/01.rti.0000203937.82276.ce](https://doi.org/10.1097/01.rti.0000203937.82276.ce).
- [169] C. Esther, M. Muhlebach, C. Ehre, *et al.*, “Mucus accumulation in the lungs precedes structural changes and infection in children with cystic fibrosis,” *Science Translational Medicine*, vol. 11, no. 486, eaav3488, 2019. DOI: [10.1126/scitranslmed.aav3488](https://doi.org/10.1126/scitranslmed.aav3488).
- [170] L. Tepper, D. Caudri, A. Rovira, H. Tiddens, and M. de Bruijne, “The development of bronchiectasis on chest computed tomography in children with cystic fibrosis: Can pre-stages be identified?” *European Radiology*, vol. 26, no. 12, pp. 4563–4569, 2016. DOI: [10.1007/s00330-016-4329-z](https://doi.org/10.1007/s00330-016-4329-z).
- [171] N. Beydon, S. Davis, E. Lombardi, *et al.*, “An official american thoracic society/european respiratory society statement: Pulmonary function testing in preschool children,” *American Journal of Respiratory and Critical Care Medicine*, vol. 175, no. 12, pp. 1304–1345, 2007. DOI: [10.1164/rccm.200605-642ST](https://doi.org/10.1164/rccm.200605-642ST).
- [172] P. Quanjer, S. Stanojevic, T. Cole, *et al.*, “Multi-ethnic reference values for spirometry for the 3–95-yr age range: The global lung function 2012 equations,” *European Respiratory Journal*, vol. 40, no. 6, pp. 1324–1343, 2012. DOI: [10.1183/09031936.00080312](https://doi.org/10.1183/09031936.00080312).

- [173] P. Sly, S. Brennan, C. Gangell, N. de Klerk, C. Murray, L. Mott, S. Stick, P. Robinson, C. Robertson, and S. Ranganathan, "Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening," *American Journal of Respiratory and Critical Care Medicine*, vol. 180, no. 2, pp. 146–152, 2009. DOI: 10.1164/rccm.200901-00690C.
- [174] A. Chang, A. Bush, and K. Grimwood, "Bronchiectasis in children: Diagnosis and treatment," *The Lancet*, vol. 392, no. 10150, pp. 866–879, 2018. DOI: [https://doi.org/10.1016/S0140-6736\(18\)31554-X](https://doi.org/10.1016/S0140-6736(18)31554-X).
- [175] L. Tepper, E. Utens, D. Caudri, A. Bos, K. Gonzalez-Graniel, H. Duivenvoorden, E. van der Wiel, A. Quittner, and H. Tiddens, "Impact of bronchiectasis and trapped air on quality of life and exacerbations in cystic fibrosis," *European Respiratory Journal*, vol. 42, no. 2, pp. 371–379, 2013. DOI: 10.1183/09031936.00137612.
- [176] D. Hansell, A. Bankier, H. MacMahon, T. McLoud, N. Müller, and J. Remy, "Fleischner society: Glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, 2008. DOI: 10.1148/radiol.2462070712.
- [177] J. Meerburg, G. Veerman, S. Aliberti, and H. Tiddens, "Diagnosis and quantification of bronchiectasis using computed tomography or magnetic resonance imaging: A systematic review," *Respiratory Medicine*, vol. 170, p. 105954, 2020. DOI: <https://doi.org/10.1016/j.rmed.2020.105954>.
- [178] S. Lonni, J. Chalmers, P. Goeminne, *et al.*, "Etiology of non-cystic fibrosis bronchiectasis in adults and its correlation to disease severity," *Annals of the American Thoracic Society*, vol. 12, no. 12, pp. 1764–1770, 2015. DOI: 10.1513/AnnalsATS.201507-4720C.
- [179] R. Chandrasekaran, M. Mac Aogáin, J. Chalmers, S. Elborn, and S. Chotirmall, "Geographic variation in the aetiology, epidemiology and microbiology of bronchiectasis," *BMC Pulmonary Medicine*, vol. 18, no. 1, p. 83, 2018. DOI: 10.1186/s12890-018-0638-0.
- [180] D. Araújo, M. Shteinberg, S. Aliberti, *et al.*, "The independent contribution of pseudomonas aeruginosa infection to long-term clinical outcomes in bronchiectasis," *European Respiratory Journal*, vol. 51, no. 2, p. 1701953, 2018. DOI: 10.1183/13993003.01953-2017.
- [181] E. Polverino, P. Goeminne, M. McDonnell, *et al.*, "European respiratory society guidelines for the management of adult bronchiectasis," *European Respiratory Journal*, vol. 50, no. 3, p. 1700629, 2017. DOI: 10.1183/13993003.00629-2017.
- [182] J. Meerburg, I. Hartmann, S. Goldacker, U. Baumann, A. Uhlmann, E. Andrinopoulou, M. Kemner van de Corput, K. Warnatz, and H. Tiddens, "Analysis of granulomatous lymphocytic interstitial lung disease using two scoring systems for computed tomography scans - a retrospective cohort study," *Frontiers in Immunology*, vol. 11, p. 589148, 2020. DOI: 10.3389/fimmu.2020.589148.

- [183] J. Chalmers, B. McHugh, C. Doherty, M. Smith, J. Govan, D. Kilpatrick, and A. Hill, "Mannose-binding lectin deficiency and disease severity in non-cystic fibrosis bronchiectasis: A prospective study," *The Lancet Respiratory Medicine*, vol. 1, no. 3, pp. 224–232, 2013. DOI: [https://doi.org/10.1016/S2213-2600\(13\)70001-8](https://doi.org/10.1016/S2213-2600(13)70001-8).
- [184] M. Loebinger, E. Polverino, J. Chalmers, *et al.*, "Efficacy and safety of TOBI podhaler in pseudomonas aeruginosa-infected bronchiectasis patients: iBEST study," *European Respiratory Journal*, vol. 57, no. 1, p. 2001451, 2021. DOI: [10.1183/13993003.01451-2020](https://doi.org/10.1183/13993003.01451-2020).
- [185] M. Loebinger, E. Polverino, F. Blasi, *et al.*, "Efficacy and safety of tobramycin inhalation powder in bronchiectasis patients with p. aeruginosa infection: Design of a dose-finding study (iBEST-1)," *Pulmonary Pharmacology & Therapeutics*, vol. 58, p. 101834, 2019. DOI: <https://doi.org/10.1016/j.pupt.2019.101834>.
- [186] A. Quittner, K. Marciel, M. Salathe, *et al.*, "A preliminary quality of life questionnaire-bronchiectasis: A patient-reported outcome measure for bronchiectasis," *Chest*, vol. 146, no. 2, pp. 437–448, 2014. DOI: <https://doi.org/10.1378/chest.13-1891>.
- [187] C. Wainwright, S. Vidmar, D. Armstrong, *et al.*, "Effect of bronchoalveolar lavage-directed therapy on pseudomonas aeruginosa infection and structural lung injury in children with cystic fibrosis: A randomized trial," *JAMA*, vol. 306, no. 2, pp. 163–171, 2011. DOI: [10.1001/jama.2011.954](https://doi.org/10.1001/jama.2011.954).
- [188] E. van Mastrigt, E. Kakar, P. Ciet, *et al.*, "Structural and functional ventilatory impairment in infants with severe bronchopulmonary dysplasia," *Pediatric Pulmonology*, vol. 52, no. 8, pp. 1029–1037, 2017. DOI: <https://doi.org/10.1002/ppul.23696>.
- [189] S. Hermelijn, O. Dragt, J. Bosch, A. Hijkoop, L. Riera, P. Ciet, R. Wijnen, J. Schnater, and H. Tiddens, "Congenital lung abnormality quantification by computed tomography: The CLAQ method," *Pediatric pulmonology*, vol. 55, no. 11, pp. 3152–3161, 2020. DOI: [10.1002/ppul.25032](https://doi.org/10.1002/ppul.25032).
- [190] T. Koo and M. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol. 15, no. 2, pp. 155–163, 2016. DOI: [10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012).
- [191] J. Chalmers, P. Goeminne, S. Aliberti, *et al.*, "The bronchiectasis severity index. an international derivation and validation study," *American Journal of Respiratory and Critical Care Medicine*, vol. 189, no. 5, pp. 576–585, 2013. DOI: [10.1164/rccm.201309-15750C](https://doi.org/10.1164/rccm.201309-15750C).
- [192] J. Chalmers, S. Aliberti, E. Polverino, *et al.*, "The EMBARC european bronchiectasis registry: Protocol for an international observational study," *ERJ Open Research*, vol. 2, no. 1, pp. 00081–2015, 2016. DOI: [10.1183/23120541.00081-2015](https://doi.org/10.1183/23120541.00081-2015).

- [193] F. Marques, F. Dubost, M. Kemner van de Corput, H. Tiddens, and M. de Bruijne, "Quantification of lung abnormalities in cystic fibrosis using deep networks," in *Medical Imaging 2018: Image Processing*, vol. 10574, SPIE, 2018, pp. 365–371. DOI: 10.1117/12.2292188.
- [194] R. Brown, "Mechanisms of limited airway dimension with lung inflation," *Pulmonary Pharmacology & Therapeutics*, vol. 20, no. 2, pp. 118–125, 2007. DOI: <https://doi.org/10.1016/j.pupt.2006.05.002>.
- [195] D. Reiff, A. Wells, D. Carr, P. Cole, and D. Hansell, "CT findings in bronchiectasis: Limited value in distinguishing between idiopathic and specific types," *American Journal of Roentgenology*, vol. 165, no. 2, pp. 261–267, 1995. DOI: 10.2214/ajr.165.2.7618537.
- [196] P. Bedi, J. Chalmers, P. Goeminne, *et al.*, "The BRICS (bronchiectasis radiologically indexed CT score): A multicenter study score for use in idiopathic and postinfective bronchiectasis," *Chest*, vol. 153, no. 5, pp. 1177–1186, 2018. DOI: <https://doi.org/10.1016/j.chest.2017.11.033>.
- [197] M. Martínez-García, J. de Gracia, M. Vendrell Relat, R. Girón, L. Máiz Carro, D. de la Rosa Carrillo, and C. Oliveira, "Multidimensional approach to non-cystic fibrosis bronchiectasis: The FACED score," *European Respiratory Journal*, vol. 43, no. 5, pp. 1357–1367, 2014. DOI: 10.1183/09031936.00026313.
- [198] C. Oliveira, G. Oliveira, F. Espildora, R. Giron, G. Muñoz, A. Quittner, and M. Martinez-Garcia, "Validation of a quality of life questionnaire for bronchiectasis: Psychometric analyses of the spanish QOL-B-V3.0," *Quality of Life Research*, vol. 23, no. 4, pp. 1279–1292, 2014. DOI: 10.1007/s11136-013-0560-0.
- [199] A. Hill, C. Haworth, S. Aliberti, *et al.*, "Pulmonary exacerbation in adults with bronchiectasis: A consensus definition for clinical research," *European Respiratory Journal*, vol. 49, no. 6, p. 1700051, 2017. DOI: 10.1183/13993003.00051-2017.
- [200] O. Weinheimer, M. Wielpütz, P. Konietzke, C. Heussel, H. Kauczor, C. Brochhausen, D. Hollemann, D. Savage, C. Galban, and T. Robinson, "Fully automated lobe-based airway taper index calculation in a low dose MDCT CF study over 4 time-points," in *Medical Imaging 2017: Image Processing*, vol. 10133, SPIE, 2017, pp. 242–250. DOI: 10.1117/12.2254387.
- [201] H. Tiddens, S. Stick, and S. Davis, "Multi-modality monitoring of cystic fibrosis lung disease: The role of chest computed tomography," *Paediatric Respiratory Reviews*, vol. 15, no. 1, pp. 92–97, 2014. DOI: <https://doi.org/10.1016/j.prrv.2013.05.003>.
- [202] H. Tiddens, J. Meerburg, M. van der Eerden, and P. Ciet, "The radiological diagnosis of bronchiectasis: What's in a name?" *European Respiratory Review*, vol. 29, no. 156, p. 190120, 2020. DOI: 10.1183/16000617.0120-2019.
- [203] J. Wichmann, M. Willemink, and C. De Cecco, "Artificial intelligence and machine learning in radiology: Current state and considerations for routine clinical implementation," *Investigative Radiology*, vol. 55, no. 9, pp. 619–627, 2020. DOI: 10.1097/RLI.0000000000000673.

- [204] K. Lloyd, *Bias amplification in artificial intelligence systems*, 2018. arXiv: 1809.07842.
- [205] G. Kaissis, M. Makowski, D. Rückert, and R. Braren, “Secure, privacy-preserving and federated machine learning in medical imaging,” *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020. DOI: 10.1038/s42256-020-0186-1.
- [206] T. Weikert, J. Cyriac, S. Yang, I. Nesic, V. Parmar, and B. Stieltjes, “A practical guide to artificial intelligence-based image analysis in radiology,” *Investigative Radiology*, vol. 55, no. 1, pp. 1–7, 2020. DOI: 10.1097/RLI.0000000000000600.
- [207] M. Willeminck, W. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L. Folio, R. Summers, D. Rubin, and M. Lungren, “Preparing medical imaging data for machine learning,” *Radiology*, vol. 295, no. 1, pp. 4–15, 2020. DOI: 10.1148/radiol.2020192224.
- [208] E. Heim, T. Roß, A. Seitel, *et al.*, “Large-scale medical image annotation with crowd-powered algorithms,” *Journal of Medical Imaging*, vol. 5, no. 3, p. 034002, 2018. DOI: 10.1117/1.JMI.5.3.034002.
- [209] V. Cheplygina, A. Perez-Rovira, W. Kuo, H. Tiddens, and M. de Bruijne, “Crowdsourcing airway annotations in chest computed tomography images,” *PLOS ONE*, vol. 16, no. 4, p. e0249580, 2021. DOI: 10.1371/journal.pone.0249580.
- [210] T. Cook, “The importance of imaging informatics and informaticists in the implementation of AI,” *Academic Radiology*, vol. 27, no. 1, pp. 113–116, 2020. DOI: <https://doi.org/10.1016/j.acra.2019.10.002>.
- [211] S. Finlayson, J. Bowers, J. Ito, J. Zittrain, A. Beam, and I. Kohane, “Adversarial attacks on medical machine learning,” *Science*, vol. 363, no. 6433, pp. 1287–1289, 2019. DOI: 10.1126/science.aaw4399.
- [212] G. Bortsova, C. Gonzalez-Gonzalo, S. Wetstein, *et al.*, “Adversarial attack vulnerability of medical image analysis systems: Unexplored factors,” *Medical Image Analysis*, vol. 73, p. 102141, 2021. DOI: <https://doi.org/10.1016/j.media.2021.102141>.
- [213] X. Yuan, P. He, Q. Zhu, and X. Li, “Adversarial examples: Attacks and defenses for deep learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2805–2824, 2019. DOI: 10.1109/TNNLS.2018.2886017.
- [214] H. Harvey and V. Gowda, “How the FDA regulates AI,” *Academic Radiology*, vol. 27, no. 1, pp. 58–61, 2020. DOI: <https://doi.org/10.1016/j.acra.2019.09.017>.
- [215] *Fda. artificial intelligence and machine learning in software as a medical device*, "<https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>", Accessed: 2022-04-01, 2020.
- [216] *Acr-dsi. fda cleared ai algorithms*, "<https://www.acrdsi.org/DSI-Services/FDA-Cleared-AI-Algorithms>", Accessed: 2022-04-01, 2019.

- [217] M. Huisman, E. Ranschaert, W. Parker, *et al.*, “An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: Fear of replacement, knowledge, and attitude,” *European Radiology*, vol. 31, no. 9, pp. 7058–7066, 2021. DOI: 10.1007/s00330-021-07781-5.
- [218] H. Cui, X. Liu, and N. Huang, “Pulmonary vessel segmentation based on orthogonal fused U-Net++ of chest CT images,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, 2019, pp. 293–300. DOI: 10.1007/978-3-030-32226-7_33.
- [219] J. Nam, J. Witanto, S. Park, S. Yoo, J. Goo, and S. Yoon, “Automatic pulmonary vessel segmentation on noncontrast chest CT: Deep learning algorithm developed using spatiotemporally matched virtual noncontrast images and low-keV contrast-enhanced vessel maps,” *European Radiology*, vol. 31, pp. 9012–9021, 2021. DOI: 10.1007/s00330-021-08036-z.
- [220] V. Cheplygina, M. de Bruijne, and J. Pluim, “Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,” *Medical Image Analysis*, vol. 54, pp. 280–296, 2019. DOI: <https://doi.org/10.1016/j.media.2019.03.009>.

